



Mariana Pedro Batista Pinto Baptista

Licenciatura em Matemática Aplicada à Economia e à Gestão

Fatores Determinantes na Fixação de Taxas de Juro em Novas Operações de Empréstimos Bancários a Sociedades Não Financeiras

Relatório de Estágio para obtenção do Grau de Mestre em
Matemática e Aplicações
Ramo de Atuariado, Estatística e Investigação Operacional

Orientador: Filipe José Gonçalves Pereira Marques,
Professor Auxiliar, Faculdade de Ciências e
Tecnologia da Universidade Nova de Lisboa
Co-orientador: Homero Alexandre Martins Gonçalves,
Coordenador da Área de Estatísticas Monetárias
Financeiras e Centralização das Responsabilidades
de Crédito, Banco de Portugal

Júri

Presidente: Marta Cristina Vieira Faias Mateus
Arguente: Pedro Alexandre da Rosa Corte Real
Vogal: Filipe José Gonçalves Pereira Marques



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro, 2016

Fatores Determinantes na Fixação de Taxas de Juro em Novas Operações de Empréstimos Bancários a Sociedades Não Financeiras

Copyright © Mariana Pedro Batista Pinto Baptista, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

AGRADECIMENTOS

A elaboração do presente trabalho só foi possível devido aos conhecimentos transmitidos, a todo o apoio, encorajamento e disponibilidade do meu orientador Professor Doutor Filipe Marques, a quem agradeço muito o contributo dado. Agradeço igualmente à Professora Doutora Marta Faias e à Sara Igrejas pelo tempo disponibilizado, simpatia e ajuda neste trabalho.

O meu agradecimento sincero ao co-orientador desta dissertação, Homero Gonçalves do Banco de Portugal, pelo profissionalismo, experiência, apoio e dedicação para comigo e com o trabalho que desenvolvi.

Ao Banco de Portugal agradeço a possibilidade de ter realizado a dissertação de Mestrado em regime de estágio curricular, assim como agradeço também, todos os dados disponibilizados, sem os quais não seria possível a elaboração do presente trabalho.

Um enorme obrigado a cada um dos elementos do Núcleo de Avaliação de Crédito, por todos os conhecimentos transmitidos, pela enorme prestabilidade, profissionalismo e companheirismo que foram essenciais durante os meus 6 meses de estágio. Ao Afonso Nunes do Núcleo de Estatísticas Monetárias e Financeiras pela disponibilidade, ajuda e sugestões dadas. À Sónia Mota e ao André Guerreiro do Núcleo de Estatísticas das Administrações Públicas, pela disponibilidade, apoio e simpatia com que me receberam e acompanharam. A todos os restantes colegas obrigada pela simpatia e incentivo para comigo.

À Andreia por toda a ajuda e apoio dados ao longo destes dois anos e por estar sempre disponível.

À Natacha pela amizade, pelos desabafos, conselhos, encorajamento e incentivo mútuo que tivemos não só durante este período intensivo de conclusão do Mestrado, como durante os últimos anos.

A todos os meus amigos e família pela amizade, apoio e companheirismo durante estes dois anos. Aos meus pais e à minha irmã um obrigado muito especial por todo o carinho, encorajamento e confiança que me deram durante o meu percurso académico e especial compreensão nas alturas mais complicadas.

Por fim, um agradecimento muito especial ao Francisco, por todo o apoio, pela tranquilidade transmitida, pelo amor e por estar sempre ao meu lado.

RESUMO

O objetivo deste trabalho prende-se com a identificação dos fatores determinantes na fixação das taxas de juro de novas operações de empréstimos bancários a sociedades não financeiras.

Em 2012, foi implementado pelo Banco de Portugal, um requisito estatístico relativo ao reporte individual das taxas de juro. Através deste requisito, o Banco de Portugal recebe, por parte das instituições financeiras, o reporte com a informação de todas as novas operações efetuadas. Deste modo, a informação disponível internamente pelo Banco de Portugal é enorme, informação essa com bastante interesse e utilidade para a realização de estudos e análises mais aprofundadas relativamente à temática das novas operações.

É recorrendo a estas bases de micro dados que se pretende encontrar os fatores que mais influenciam a fixação das taxas de juro bancárias das novas operações, identificando assim, os seus fatores agravantes e desagравantes.

Foi realizada uma análise econométrica, com o intuito de alcançar o objetivo enunciado, avaliando assim a influência conjunta das variáveis consideradas na explicação da fixação das taxas de juro. Os modelos de regressão abordados são o modelo de regressão linear múltipla e o modelo de regressão beta, sendo possível, a partir destes, tirar conclusões para dar resposta ao problema proposto.

Posteriormente foi efetuada uma análise com o objetivo de proceder à determinação da importância relativa das variáveis no que diz respeito ao modelo de regressão linear, utilizando diferentes métodos de medição da referida importância.

Devido ao facto da amostra utilizada ter uma dimensão bastante elevada, foi efetuada uma avaliação do impacto que a dimensão da amostra tem nos resultados estatísticos provenientes dos modelos de regressão analisados.

Palavras-chave: Taxas de Juro; Novas Operações; Empréstimos; Sociedades Não Financeiras; Modelo de Regressão Linear Múltipla; Modelo de Regressão Beta; Importância Relativa; Dimensão da Amostra.

ABSTRACT

The objective of this study is to identify the determinant factors in the setting of interest rates of new operations of bank loans to non-financial corporations.

In 2012, a requirement of statistics concerning the report of individual interest rates was implemented by the Bank of Portugal. Through this requirement, the Bank of Portugal receives, by financial institutions, the report with the information of all new operations carried out. In this way, the information available internally by the Bank of Portugal is huge, with great interest and usefulness to perform studies and further analysis regarding the issue of new operations.

Through these databases of micro data it is possible to find the factors that most influence the setting of interest rates accounts of new operations, thereby identifying their aggravating factors and desaggravantes.

An econometric analysis was performed with the aim of achieving the goal enunciated, thus evaluating the joint influence of the variables considered in the explanation of the setting of interest rates. The regression models discussed are the multiple linear regression and the beta regression, being possible, from these, to draw conclusions in response to the proposed problem.

An analysis, was later carried out, with the objective to determine the relative importance of variables with respect to the linear regression model, using different methods of measurement of such importance.

Due to the fact that the sample used has a quite high dimension, an evaluation was performed of the impact that the size of the sample has in the statistical results from the regression models examined.

Keywords: Interest Rate; New Operations; Loans; Non-financial Corporations; Multiple Linear Regression Model; Beta Regression Model; Relative Importance; Sample Size.

ÍNDICE

Lista de Figuras	xv
Lista de Tabelas	xvii
1 Introdução	1
2 Enquadramento Teórico	5
2.1 Modelo de Regressão Linear Múltipla	5
2.1.1 Transformação de Box-Cox	9
2.1.2 Importância Relativa dos Regressores	9
2.2 Modelo de Regressão Beta	12
2.3 Método de Seleção de Variáveis	17
2.3.1 Seleção Forward	17
2.3.2 Eliminação Backward	18
2.3.3 Método Stepwise	18
2.4 Análise da Variação dos Resultados em Função da Dimensão das Amostras	18
3 Modelação dos Dados	23
3.1 Informação Disponibilizada pelo Banco de Portugal	23
3.2 Caracterização dos Dados	25
3.3 Análise das Variáveis Explicativas	31
3.3.1 Características do Empréstimo	32
3.3.2 Características da Empresa	39
3.4 Análise Univariada	46
3.5 Análise Multivariada	48
3.5.1 Modelo de Regressão Linear Múltipla	51
3.5.2 Modelo de Regressão Beta	56
3.5.3 Comparação dos Modelos	61
4 Análise de Resultados	63
4.1 Importância Relativa dos Regressores	63
4.2 Comparação das Instituições Financeiras Nacionais e Estrangeiras	67
5 Análise da Variação dos Resultados em Função da Dimensão das Amostras	71

ÍNDICE

5.1	Modelo de Regressão Linear Múltipla	71
5.2	Modelo de Regressão Beta	74
6	Considerações Finais	79
	Bibliografia	81

LISTA DE FIGURAS

2.1	Densidades da distribuição Beta para diferentes combinações de (μ, ϕ) (Fonte: Ferrari e Cribari-Neto (2004))	14
3.1	Taxas referentes a empréstimos e depósitos de SNF's e particulares	24
3.2	Evolução da TAA de empréstimos a sociedades não financeiras	25
3.3	Caracterização dos dados observados	26
3.4	Evolução da TAA média	27
3.5	Histograma da TAA	27
3.6	Evolução das taxas consoante o montante associado ao empréstimo	33
3.7	Evolução das taxas consoante o prazo contratual associado ao empréstimo	34
3.8	Evolução das taxas consoante o prazo de fixação inicial da taxa do empréstimo	35
3.9	Evolução das taxas consoante o tipo de negociação do empréstimo	36
3.10	Evolução das taxas com a distinção de empréstimos com e sem colateral	37
3.11	Evolução das taxas consoante os bancos pertencentes ao Top 5	38
3.12	Evolução das taxas com a distinção de IF's nacionais e estrangeiras	38
3.13	Evolução das taxas consoante a dimensão da empresa	40
3.14	Evolução das taxas consoante o setor de atividade da empresa	41
3.15	Evolução das taxas consoante a classe de rating da empresa	42
3.16	Evolução das taxas com a distinção da empresa ter ou não crédito vencido	43
3.17	Evolução das taxas consoante a característica PME líder	43
3.18	Evolução das taxas consoante a característica relativa às exportações	44
3.19	Evolução das taxas consoante a característica EEC	44
3.20	Evolução das taxas consoante o número de instituições financeiras	45
3.21	Evolução das taxas consoante o número de novas operações	46
3.22	Análise do ajustamento das distribuições teóricas à distribuição empírica	49
3.23	Determinação do valor de λ	51
3.24	Aplicação da transformação de Box-Cox à variável resposta	52
3.25	Resíduos de Pearson em função dos valores estimados e das observações no modelo de regressão linear múltipla	56
3.26	Distâncias de Cook no modelo de regressão linear múltipla	57
3.27	Resíduos de Pearson e resíduos padronizados ponderados aplicados ao modelo de regressão beta	60

3.28	Distâncias de Cook aplicadas ao modelo de regressão beta	60
3.29	Comparação das densidades empírica e teóricas relativas às taxas de juro . .	62
4.1	Comparação dos métodos First, Last, LMG e PMVD na avaliação da importância das variáveis em estudo	64
4.2	Densidade das taxas de juro comparando IF's nacionais e estrangeiras	68
4.3	Importância dos regressores comparando IF's nacionais e estrangeiras	69
5.1	Coefficiente, intervalo de confiança a 95% e p -value associados à variável Montante através do modelo de regressão linear	72
5.2	Análise do p -value do teste F	73
5.3	Análise do R^2 e do \bar{R}^2	74
5.4	Coefficiente, intervalo de confiança a 95% e p -value associados à variável Montante através do modelo de regressão beta	75
5.5	Análise do p -value do teste de razão de verossimilhanças	76
5.6	Análise do valor do pseudo R^2	76

LISTA DE TABELAS

3.1	Secções CAE	29
3.2	Classes de Rating	30
3.3	Variáveis explicativas relativas às características do empréstimo	32
3.4	Variáveis explicativas relativas às características da empresa	39
3.5	Análise univariada dos dois modelos em estudo	48
3.6	Variáveis incluídas nos modelos	50
3.7	Fatores padrão	52
3.8	Resultados do Modelo de Regressão Linear Múltipla	54
3.9	Escolha da função de ligação	57
3.10	Resultados do Modelo de Regressão Beta	58
3.11	Comparação das estimativas dos coeficientes pelos dois modelos de regressão	61
4.1	Características do empréstimo relativo à empresa ABC	65
4.2	Variações nas taxas de juro	66
4.3	Variações nas taxas de juro comparando IF's nacionais e estrangeiras	68

INTRODUÇÃO

A recolha e elaboração de estatísticas financeiras e monetárias é da competência do Banco de Portugal, sendo esta função inserida no seu contributo para o sistema estatístico nacional. A necessidade do Banco de Portugal em realizar estudos e análises acerca da economia portuguesa e do seu sistema financeira, implica uma recolha e compilação detalhada de informação. Por estes motivos, em 2012, foi introduzido um conjunto de novos requisitos estatísticos, entre eles, o reporte de informação individual das taxas de juro sobre novas operações de empréstimos concedidos a sociedades não financeiras. Deste modo, as instituições financeiras reportam ao Banco de Portugal o conjunto de todas as novas operações de empréstimos, sendo criada uma base de micro dados.

As bases de micro dados resultam de uma recolha intensiva de dados, que só é possível devido à facilidade de acesso à informação verificada nos dias de hoje. Deste modo, a informação disponível caracteriza-se por ser bastante mais detalhada o que permite a obtenção de um variado conjunto de análises e estudos. Como referido em Antunes et al. (2013), são várias as potencialidades dos micro dados. Esta tipologia de dados permite efetuar agregações de variáveis, trabalha-las da forma mais conveniente para cada tipo de estudo a realizar e possibilitando ainda a construção de novas variáveis com base nas já existentes. Deste modo verifica-se uma melhoria da qualidade dos dados disponíveis assim como da capacidade de resposta a determinados pedidos de informação específica.

O reporte individual das taxas de juro permite caracterizar não só o tipo de empréstimos realizados, mas também os mutuários e mutuantes. Isto acontece porque a informação contida nas bases de micro dados permite efetuar um cruzamento com as informações das restantes bases de dados disponíveis no Banco de Portugal. A utilização simultânea de informações de diferentes bases de dados, permite complementar o conjunto de características disponíveis, enriquecendo a realização de estudos e análises.

O objetivo do presente estudo prende-se com a identificação do conjunto de características determinantes na fixação das taxas de juro bancárias de novos empréstimos concedidos a empresas, utilizando como ponto de partida as referidas bases de micro dados. Serão utilizadas variáveis não financeiras, caracterizadoras do empréstimo em causa, da respetiva empresa e da instituição financeira que concede o empréstimo, com o intuito de perceber quais os fatores que mais influenciam e diferenciam a atribuição das taxas de juro a novos empréstimos. Em Martinho e Antunes (2012a) e Martinho e Antunes (2012b) são efetuados estudos econométricos com base nas características das empresas, características essas contempladas também no presente trabalho.

Deste modo, torna-se interessante efetuar uma análise econométrica com o intuito de determinar o conjunto de fatores mais influentes na fixação das taxas de juro bancárias, tanto características do empréstimo como da empresa à qual foi concedido o referido empréstimo.

No presente trabalho, o Capítulo 2 é dedicado ao enquadramento teórico das temáticas que serão apresentadas posteriormente. Os Capítulos 3, 4 e 5 apresentam detalhadamente toda a aplicação prática realizada ao longo deste trabalho e, por fim, o Capítulo 6 invoca algumas considerações finais relativamente ao trabalho realizado.

O Capítulo 2, tal como referido, apresenta um enquadramento teórico das temáticas abordadas ao longo do presente trabalho. Serão apresentados dois modelos de regressão, o modelo linear e o modelo beta, que, tal como será verificado, ajustam-se ao comportamento da variável resposta que se pretende estudar. No que diz respeito ao modelo de regressão linear múltipla, será apresentado um conjunto de métodos com o intuito de avaliar a importância relativa dos regressores, que será posteriormente aplicado no capítulo 4. Será ainda abordada, em termos teóricos, uma metodologia de avaliação de resultados estatísticos resultantes de amostras de grandes dimensões.

Os modelos de regressão apresentados no enquadramento teórico serão aplicados no Capítulo 3. Neste capítulo, será efetuada uma apresentação e caracterização das variáveis explicativas e, posteriormente, uma análise univariada e multivariada, com o intuito entender e tentar explicar o comportamento das taxas de juro de novas operações de empréstimos a sociedades não financeiras. Serão identificados os fatores que originam um agravamento ou um desagravamento das taxas de juro.

Como mencionado anteriormente, será realizada no Capítulo 4 uma avaliação da importância relativa de cada um dos regressores utilizando algumas das medidas desenvolvidas no Capítulo 2. Através desta abordagem será possível avaliar e quantificar o impacto de cada um dos regressos na fixação das taxas de juro em análise.

No Capítulo 5 realizar-se-á uma abordagem relativa ao impacto da dimensão das amostras nos resultados estatísticos obtidos. Serão realizadas reamostragens de amostras mais reduzidas, com a finalidade de avaliar o comportamento dos coeficientes estimados e dos p -values associados, assim como de algumas medidas de ajustamento dos modelos, em função da dimensão da amostra.

Por fim, será realizada uma avaliação global dos resultados obtidos através deste trabalho e apresentar-se-á, ainda, um conjunto de considerações finais contemplando estudos interessantes para trabalhos futuros.

ENQUADRAMENTO TEÓRICO

2.1 Modelo de Regressão Linear Múltipla

Uma análise de regressão estuda a relação entre uma variável resposta e uma ou mais variáveis explicativas, também denominadas de regressores. Esta relação é explicada utilizando modelos matemáticos, conhecidos como modelos de regressão.

O modelo de regressão mais simples é o modelo de regressão linear, que pressupõe uma relação linear entre a variável resposta e as variáveis explicativas. O modelo de regressão linear, que considera uma variável resposta e apenas uma variável explicativa, é denominado de modelo de regressão linear simples. Caso sejam consideradas duas ou mais variáveis explicativas, tem-se um modelo de regressão linear múltipla.

Mais especificamente, o modelo de regressão linear múltipla descreve uma relação entre p regressores, x_j , $j = 1, \dots, p$ e a variável resposta Y .

Tem-se

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

onde os parâmetros β_j são os coeficientes de regressão e ε é o erro aleatório associado ao modelo.

Considerando uma amostra de dimensão n tem-se

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n$$

onde Y_i é a variável resposta, x_{i1}, \dots, x_{ip} os regressores, ou variáveis explicativas, β_1, \dots, β_p são os coeficientes de regressão e ε_i é o erro aleatório associado ao modelo.

Os pressupostos deste modelo são

$$E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2, \quad Cov(\varepsilon_i, \varepsilon_k) = 0, \quad i \neq k \quad e \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

O modelo de regressão linear múltipla, pode também ser escrito em notação matricial

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \Leftrightarrow \underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

em que

- \underline{Y} é o vetor das observações da variável resposta;
- X é a matriz das observações das variáveis explicativas;
- $\underline{\beta}$ é o vetor dos coeficientes de regressão;
- $\underline{\varepsilon}$ é o vetor dos resíduos.

Com base nos dados disponíveis é possível estimar $\beta_0, \beta_1, \dots, \beta_p$. Estes parâmetros aos serem substituídos pelas suas estimativas $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ permitem obter a equação de regressão estimada, dada por

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}, \quad i = 1, \dots, n$$

A estimação de $\underline{\beta}$ é feita através do método dos Mínimos Quadrados, cujo objetivo é a minimização do quadrado dos resíduos, ou seja, a minimização do quadrado da distância entre os valores observados e os valores estimados. Deste modo, a função que se pretende minimizar é a seguinte

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2.$$

O estimador de $\underline{\beta}$ resultante da aplicação do método dos Mínimos Quadrados é

$$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (X^T X)^{-1} X^T Y$$

Através do estimador dos Mínimos Quadrados de $\underline{\beta}$, obtém-se as estimativas para os coeficientes de regressão do modelo, e, deste modo, tem-se os valores ajustados para a variável resposta, ou seja

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}, \quad i = 1, \dots, n$$

Os resíduos dos mínimos quadrados são dados pela expressão

$$\hat{e}_i = y_i - \hat{y}_i, \quad i = 1, \dots, n$$

A análise de variância é uma análise bastante importante na temática dos modelos de regressão linear múltipla. Esta análise baseia-se na decomposição da soma dos quadrados totais, SQT , na soma dos quadrados explicada, SQE , e na soma dos quadrados dos resíduos, SQR . A soma dos quadrados explicada corresponde à variação explicada da variável resposta, enquanto que a soma dos quadrados dos resíduos corresponde à variação não explicada. A soma dos quadrados totais corresponde à variação total da variável resposta. Tem-se que

- $SQT = \sum_{i=1}^n (y_i - \bar{y})^2$
- $SQE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$
- $SQR = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (\hat{y}_i - y_i)^2$

Verifica-se, deste modo, a seguinte relação

$$SQT = SQE + SQR \Leftrightarrow \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

A qualidade do ajuste pode ser avaliada através do coeficiente de determinação, R^2 . O coeficiente de determinação mede a proporção da variação de y explicada pelos p regressores do modelo e é dado pela expressão

$$R^2 = 1 - \frac{SQR}{SQT} = \frac{SQE}{SQT}.$$

O valor do R^2 aumenta sempre que são adicionadas variáveis explicativas ao modelo. Por outro lado, o coeficiente de determinação ajustado, \bar{R}^2 , diz respeito a uma versão modificada do R^2 que não aumenta necessariamente quando são adicionados regressores ao modelo. A sua expressão é dada por

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} \frac{SQR}{SQT} = 1 - \frac{n-1}{n-k-1} (1 - R^2)$$

Neste caso, a inclusão de regressores com pouco poder explicativo reduz o valor de \bar{R}^2 , na medida em que k aumenta uma unidade, mas, no entanto, o valor de R^2 não aumenta significativamente. O valor do coeficiente de determinação, R^2 , assim como o do coeficiente de determinação ajustado, \bar{R}^2 , variam entre 0 e 1. Quanto mais próximo de 1 for o valor de \bar{R}^2 , mais explicativo é o modelo, existindo um melhor ajustamento do modelo aos dados.

Recorrendo à estatística de Wald, é possível realizar testes à significância individual de cada um dos regressores. Usualmente, pretende-se testar $H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0$, considerando-se a seguinte estatística de teste

$$W = \frac{\hat{\beta}_j}{\widehat{se}(\hat{\beta}_j)}.$$

Pode também realizar-se um teste à significância global do modelo, com o objetivo de avaliar se a equação de regressão explica, ou não, a variação na variável resposta. Para isso, realiza-se um teste de hipóteses, em que a hipótese nula considera que os regressores não são conjuntamente significativos, ou seja, o modelo não é adequado para explicar a variável resposta em estudo. Matematicamente, este teste é definido por

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_p = 0 \quad vs \quad H_1 : \exists j : \beta_j \neq 0, j = 1, \dots, p$$

Para testar a hipótese nula, H_0 , utiliza-se a estatística de teste F , dada por

$$F = \frac{\frac{SQE}{p}}{\frac{SQR}{n-p-1}} = \frac{n-k-1}{k} \frac{R^2}{1-R^2} \sim F_{p,n-p-1}$$

onde $\frac{R^2}{1-R^2}$ diz respeito à razão entre a variação explicada e a não explicada em y .

A hipótese nula é rejeitada quando existe evidência estatística de que os regressores são conjuntamente significativos para a explicação da variável resposta, pelo que devem ser incluídos no modelo de regressão. Caso a hipótese nula não seja rejeitada, verifica-se que o conjunto de regressores do modelo contribuem pouco para a explicação da variação da variável resposta.

Existem outros testes e procedimentos interessantes, sendo um deles o teste a submodelos. Esta tipologia de testes, baseada na análise ANOVA, permite comparar dois ou mais modelos encaixados. Num modelo com p regressores, pretende testar-se se q desses regressores, $p > q$, são conjuntamente significativos para o modelo. De acordo com Blackwell (2008), considera-se SQR_r a soma do quadrado dos resíduos do modelo com restrições e SQR_{sr} a soma do quadrado dos resíduos do modelo sem restrições. A hipótese nula é da forma $H_0 : \beta_0 = \beta_1 = \dots = \beta_q = 0$ e a estatística F utilizada é bastante semelhante à enunciada anteriormente, sendo dada por

$$F = \frac{\frac{SQR_r - SQR_{sr}}{q}}{\frac{SQR_{sr}}{n-p-1}} \sim F_{q,n-p-1}.$$

Quanto maior a diferença dada por $SQR_r - SQR_{sr}$, pior é o ajustamento, uma vez que, caso as variáveis não sejam significativas, o facto de eliminá-las não afetará os resíduos.

A análise de resíduos é de extrema importância na investigação da adequabilidade do modelo de regressão. A ideia base da análise de resíduos afirma que, se o modelo for apropriado, os resíduos deverão refletir as propriedades enunciadas no início desta secção, tais como independência, variância constante e distribuição Normal.

A distância de Cook é também um método muito utilizado na avaliação do ajuste do modelo, que calcula o impacto de uma observação na estimação do parâmetro de regressão, medindo o efeito de se excluir uma determinada observação. Deste modo, é possível identificar observações altamente influentes.

2.1.1 Transformação de Box-Cox

Nem sempre a distribuição Normal se adequa aos dados em estudo, sendo possível, nesses casos, aplicar uma transformação aos dados e normalizá-los. A transformação de dados é uma prática comum existindo várias que podem ser utilizadas para aproximar a distribuição dos dados da distribuição Normal, sendo uma delas a transformação de Box-Cox, ver Osborne (2010).

A transformação de Box-Cox consiste em encontrar um λ de forma a que a variável resposta transformada se aproxime de uma distribuição Normal, reduzindo assim as assimetrias dos dados. A referida transformação, proposta por Box e Cox (1964), é dada pela expressão

$$y_i^{BC} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y_i), & \lambda = 0 \end{cases}$$

A estimação do parâmetro λ é feita através da máxima verosimilhança, ou seja, o valor atribuído a λ é o valor que maximiza a função de log-verosimilhança de y_i^{BC} em relação às observações originais. Segundo Box e Cox (1964), depois de encontrado o λ , obtém-se uma transformação adequada para a variável resposta, capaz de tornar os dados aproximadamente com distribuição Normal.

2.1.2 Importância Relativa dos Regressores

O tópico da importância relativa dos regressores em modelos de regressão tem sido investigado e aprofundado nos últimos anos, devido a um aumento no interesse deste assunto em aplicações práticas de diversos estudos e análises.

A denominada importância relativa refere-se à quantificação da contribuição individual de cada regressor para o modelo de regressão múltipla. Johnson e LeBreton (2004) definiram o conceito de importância relativa, como a contribuição que cada regressor proporciona ao R^2 , considerando não só o seu efeito direto mas também o efeito combinado com as restantes variáveis do modelo de regressão. Se os regressores não

forem correlacionados entre si, a contribuição de cada um deles corresponde ao valor de R^2 relativo à regressão univariada. Consequentemente, a soma do R^2 de todas a regressões univariadas iguala o valor do R^2 do modelo completo. No entanto, são vários os casos em que existe alguma correlação entre os regressores, e, desse modo, não é possível decompor o R^2 do modelo através do R^2 das regressões univariadas.

Tal como visto anteriormente, o coeficiente de determinação representa a proporção da variação na variável resposta explicada pelos regressores do modelo, sendo a sua expressão dada por

$$R^2 = \frac{SQE}{SQT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Em Grömping et al. (2006) são apresentados e implementados, no *software R*, seis diferentes métodos para avaliar a importância relativa em regressões lineares.

As métricas respeitantes à importância relativa dos regressores, apresentadas neste artigo, distinguem-se entre métricas simples e métricas computacionalmente intensivas. São apresentadas quatro métricas simples (*first*, *last*, *betasq* e *pratt*) e duas métricas com um esforço computacional mais elevado (*lmg* e *pmvd*).

A **métrica *first*** compara a contribuição individual de cada regressor na explicação da variável resposta, ou seja, compara o valor do R^2 das regressões univariadas dos p regressores do modelo. A métrica *first* é a única que não considera a influência de outros regressores do modelo. Esta métrica não é a mais adequada, na medida em que, nem sempre o R^2 pode ser decomposto em contribuições individuais dos regressores, fazendo com que a soma dessas contribuições não iguale o R^2 do modelo.

A **métrica *last*** é uma outra alternativa na temática da importância relativa dos regressores, que compara a contribuição adicional que cada regressor proporciona ao modelo quando incluído como último regressor. Esta métrica é caracterizada por ser muito semelhante à análise feita através de testes de significância dos regressores. Tal como se verifica na métrica *first*, também na métrica *last* as contribuições dos regressores podem não somar o valor do R^2 do modelo.

Outra abordagem à temática da importância relativa dos regressores é a **métrica *betasq*** que considera uma padronização dos coeficientes de regressão. Segundo o artigo em análise, a padronização dos coeficientes é dada por

$$\hat{\beta}_{k,standardized} = \hat{\beta}_k \frac{\sqrt{s_{kk}}}{\sqrt{s_{yy}}}$$

onde s_{kk} é a variância amostral do regressor k e s_{yy} é a variância amostral da variável resposta. A métrica *betasq* consiste então, na análise do quadrado da padronização dos

coeficientes de regressão do modelo. No entanto, não é considerada uma boa métrica na medida em que não fornece uma decomposição natural do R^2 .

A **métrica *pratt*** é a última das quatro métricas simples apresentadas em Grömping et al. (2006). Esta métrica, foi inicialmente proposta por Hoffman (1960) e baseia-se no produto do coeficiente padronizado e da correlação entre o regressor e a variável resposta. Hoffman considerou esta métrica como uma decomposição do R^2 , tendo sido uma proposta por muitos criticada, mas apoiada e defendida por Pratt em 1987. A maior crítica à métrica *pratt* é o facto da decomposição do R^2 poder considerar contribuições negativas. Deste modo, a métrica *pratt* não pode ser aplicada em todas as situações.

As quatro métricas descritas anteriormente são as denominadas métricas simples para analisar a importância relativa das variáveis em regressões lineares. As duas métricas apresentadas de seguida, *lmg* e *pmvd*, correspondem a métricas mais complexas e que exigem um maior esforço computacional. Ambas as métricas decompõem o R^2 em contribuições não negativas que somam automaticamente o coeficiente de determinação. A dificuldade de decomposição do R^2 em modelos com variáveis explicativas correlacionadas assenta no facto de cada ordenação de regressores produzir uma decomposição diferente da soma dos quadrados totais, *SQT*.

Segundo Grömping et al. (2006), as abordagens das métricas *lmg* e *pmvd* são baseadas em R^2 sequenciais. Por sequencial considera-se que os regressores são introduzidos no modelo pela ordem em que estão listados.

Sabendo que o R^2 para um modelo com um conjunto de regressores S é

$$R^2(S) = \frac{SQE}{SQT}$$

O R^2 adicional resultante da junção de um conjunto de regressores M a uma regressão com conjunto de regressores S é dado por

$$seqR^2(M|S) = R^2(M \cup S) - R^2(S)$$

Em qualquer modelo, a ordem dos regressores corresponde a uma permutação dos regressores x_1, \dots, x_p e é definido por $r = (r_1, \dots, r_p)$. Seja $S_k(r)$ o conjunto de regressores incluídos no modelo antes do regressor x_k , a parcela do R^2 relativa ao regressor x_k é dada por

$$seqR^2(\{x_k\}|S_k(r)) = R^2(\{x_k\} \cup S_k(r)) - R^2(S_k(r))$$

A **métrica *lmg***, proposta por Lindeman et al. (1980) é descrita através da expressão seguinte

$$LMG(x_k) = \frac{1}{p!} \sum_{r \text{ permutation}} seqR^2(\{x_k\}|r).$$

A métrica *lmg* propõe um uso sequencial da soma dos quadrados dos modelos que depende da ordem dos regressores no modelo. Obtém-se assim uma análise global fazendo a média de todas as ordenações dos regressores. Ou seja, esta métrica consiste no facto de se ter as médias das contribuições dos regressores para todas as possíveis ordenações.

Uma das críticas à métrica *lmg*, baseia-se no facto de que, com esta métrica, um regressor com coeficiente zero possa ter uma contribuição positiva, mas, no entanto é uma métrica recomendada por vários autores, entre os mais Grömping et al. (2006).

A métrica *pmvd* foi proposta por Feldman (2005) e pretende corrigir esta desvantagem. Esta métrica garante que um regressor cujo coeficiente estimado seja zero, terá uma importância relativa também de zero. Garante-se ainda que a contribuição converge para zero se o verdadeiro valor do coeficiente for zero. Deste modo, a métrica *pmvd* resulta também da média das ordenações dos regressores, mas neste caso, os dados dependem dos pesos das ordenações dos regressores,

$$PMVD(x_k) = \frac{1}{p!} \sum_{r \text{ permutation}} p(r) seqR^2(\{x_k\}|r),$$

onde $p(r)$ representa o peso das ordenações, sendo dado por

$$p(r) = \frac{L(r)}{\sum_{r \text{ permutation}} L(r)}$$

onde

$$L(r) = \prod_{i=1}^{p-1} seqR^2(\{x_{r_{k+1}}, \dots, x_{r_p}\}|\{x_{r_1}, \dots, x_{r_k}\})^{-1}.$$

A métrica *pmvd* pode ser interpretada como uma média ponderada relativamente às ordenações dos regressores, em que os pesos dependem dos dados.

Das métricas apresentadas, e disponíveis para aplicação no package 'relaimpo' do software R, as recomendadas pela literatura, que serão utilizadas neste trabalho, são as métricas *lmg* e *pmvd*, na medida em que são as que mais se aproximam da definição de importância relativa dos regressores, sugerida por Johnson e LeBreton (2004).

2.2 Modelo de Regressão Beta

O modelo de regressão beta baseia-se na hipótese de que a variável resposta tem uma distribuição Beta. Tal como propuseram Ferrari e Cribari-Neto (2004), as regressões beta são apropriadas para modelar variáveis resposta que assumem valores contínuos pertencentes ao intervalo (0,1), como percentagens, taxas e proporções.

A densidade da distribuição Beta é dada por

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1,$$

em que $\Gamma(\cdot)$ é a função Gama e $p, q > 0$.

A média e a variância são, respetivamente,

$$E(y) = \frac{p}{p+q} \quad \text{e} \quad \text{var}(y) = \frac{pq}{(p+q)^2(p+q+1)}.$$

O modelo de regressão beta, proposto por Ferrari e Cribari-Neto (2004), considera uma parametrização da distribuição Beta em termos da média e de um parâmetro de precisão. Deste modo, tem-se $\mu = \frac{p}{p+q}$ e $\phi = p+q$, isto é, $p = \mu\phi$ e $q = (1-\mu)\phi$, obtendo-se assim a parametrização referida

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1,$$

onde $0 < \mu < 1$ e $\phi > 0$. A média e a variância da parametrização são

$$E(y) = \mu \quad \text{e} \quad \text{var}(y) = \frac{\mu(1-\mu)}{1+\phi}$$

O parâmetro μ é, tal como foi referido, a média da variável resposta e ϕ é designado por parâmetro de precisão. Para um dado valor fixo de μ , quanto maior for o valor de ϕ , menor é a variância de y . A distribuição Beta é bastante flexível, na medida em que a sua densidade consegue tomar diferentes formas, como se pode observar na Figura 2.1.

Sejam y_1, \dots, y_n as variáveis explicativas, tal que $y_i \sim B(\mu_i, \phi)$, para $i = 1, \dots, n$. Deste modo, o modelo de regressão beta é definido por

$$g(\mu_i) = \underline{x}_i^T \underline{\beta} = \eta_i$$

onde o vetor dos parâmetros de regressão é $\underline{\beta} = (\beta_1, \dots, \beta_k)^T$, $\underline{x}_i = (x_{i1}, \dots, x_{ik})^T$ é o vetor das observações das variáveis explicativas e k os regressores do modelo. A função $g(\cdot) : (0, 1) \mapsto \mathbb{R}$ é a função de ligação do modelo, sendo estritamente crescente e duas vezes diferenciável.

Existem dois motivos principais que justificam a utilização de uma função de ligação. Um deles é o facto de ambos os membros da equação de regressão assumirem valores reais quando é aplicada uma função de ligação a μ_i . Outro motivo é a existência de uma maior flexibilidade, na medida em que é possível escolher a função de ligação que melhor se adequa ao problema em estudo.

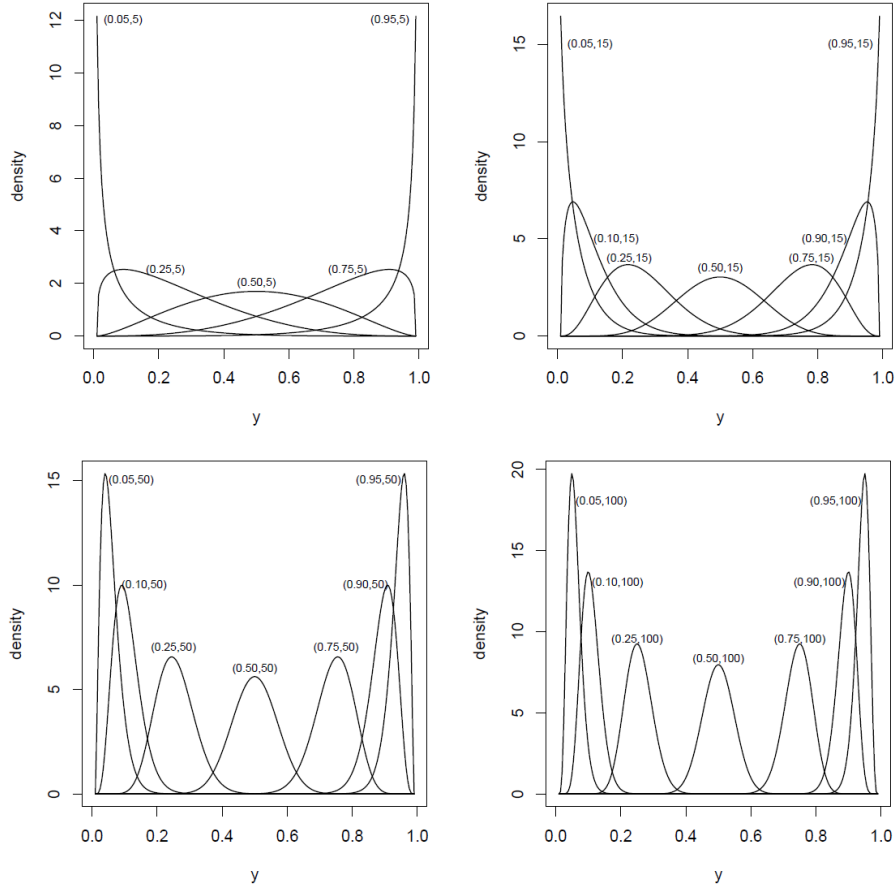


Figura 2.1: Densidades da distribuição Beta para diferentes combinações de (μ, ϕ) (Fonte: Ferrari e Cribari-Neto (2004))

As principais funções de ligação utilizadas no modelo de regressão beta são:

- Logit: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
- Probit: $g(\mu) = \Phi^{-1}(\mu)$
- Log-log complementar: $g(\mu) = \log\{-\log(1-\mu)\}$
- Log-log: $g(\mu) = -\log\{-\log(\mu)\}$
- Cauchy: $g(\mu) = \tan\{\pi(\mu - 0.5)\}$

A função de log-verosimilhança é

$$l(\beta, \phi) = \sum_{i=1}^n l_i(\mu_i, \phi)$$

onde

$$l_i(\mu_i, \phi) = \log\Gamma(\phi) - \log\Gamma(\mu_i\phi) - \log\Gamma((1-\mu_i)\phi) + (\mu_i\phi - 1)\log y_i + \{(1-\mu_i)\phi - 1\}\log(1-y_i)$$

Note-se que $\mu_i = g^{-1}(x_i^T \beta)$ é função dos parâmetros de regressão. A estimação destes parâmetros é feita com base no método da máxima verosimilhança.

Sob as condições usuais de regularidade para estimadores de máxima verosimilhança, e com base nos pressupostos subjacentes em Batista (2009), tem-se que, para grandes amostras

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \overset{a}{\sim} N_{k+1} \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K^{-1} \right)$$

onde $\hat{\beta}$ e $\hat{\phi}$ são respetivamente os estimadores de máxima verosimilhança de β e de ϕ e K a matriz de Informação de Fisher.

Apenas como curiosidade, referimos que adicionalmente ao modelo de regressão beta com dispersão constante, foi estudado e utilizado por Smithson e Verkuilen (2006), um modelo de regressão beta com dispersão variável, formalmente introduzido por Simas et al. (2010).

Contrariamente ao que se verifica no modelo de regressão beta descrito anteriormente, cujo parâmetro de dispersão é constante para todas as observações, neste caso, o parâmetro de dispersão é modelado de forma semelhante ao parâmetro μ . Assim, $y_i \sim B(\mu_i, \phi_i)$, $i = 1, \dots, n$ e, tem-se que

$$g_1(\mu_i) = \eta_{1i} = x_i^T \beta,$$

$$g_2(\phi_i) = \eta_{2i} = z_i^T \gamma,$$

onde $\beta = (\beta_1, \dots, \beta_k)^T$, $\gamma = (\gamma_1, \dots, \gamma_h)^T$, $k + h < n$ são os coeficientes de regressão das duas equações, η_{1i} e η_{2i} são os preditores lineares, e x_i e z_i são os vetores dos regressores. Tal como no modelo de regressão beta com dispersão constante, os parâmetros μ_i e ϕ_i são estimados pelo método de máxima verosimilhança.

Para avaliar a qualidade do ajustamento do modelo, uma das medidas que avalia o ajustamento global é o pseudo R^2 . O valor do pseudo R^2 , R_p^2 , corresponde ao quadrado do coeficiente de correlação entre o preditor linear e as observações da variável resposta, transformadas pela função de ligação. Uma vez que $0 < R_p^2 < 1$, verifica-se a existência de ajustamento perfeito entre $\hat{\eta}$ e $g(y)$ quando $R_p^2 = 1$.

A avaliação da qualidade do ajustamento do modelo, pode também ser analisada através da Deviance, que mede a discrepância do ajustamento, sendo dada pela expressão

$$D(y; \mu, \phi) = \sum_{i=1}^n 2(l_i(\tilde{\mu}_i, \phi) - l_i(\mu_i, \phi))$$

onde $\tilde{\mu}_i$ é a solução da equação $\frac{\partial l_i}{\partial \mu_i} = 0$.

A validação do modelo deve também ser feita com base numa análise dos resíduos, mas, tal como referido em Cribari-Neto e Zeileis (2009), os resíduos usuais não devem ser usados devido ao facto do modelo ser heterocedástico. No entanto, são vários os tipos de resíduos que podem ser analisados nos modelos de regressão beta, como por exemplo os resíduos de Pearson, também denominados de resíduos comuns padronizados (*stantardized ordinary residuals*) por Ferrari e Cribari-Neto (2004). Estes resíduos são definidos por

$$r_{P,i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{VAR}(y_i)}},$$

onde $\widehat{VAR}(y_i) = \frac{\hat{\mu}_i(1 - \hat{\mu}_i)}{(1 + \hat{\phi})}$ e $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$.

Um outro tipo de resíduo, proposto por Espinheira et al. (2008b), que permite avaliar o ajuste do modelo, é o denominado *standardized weighted residual 2*, correspondendo a resíduos padronizados ponderados, definidos por

$$r_{sw2,i} = \frac{y_i^* - \hat{\mu}_i^*}{\sqrt{\hat{v}_i(1 - h_{ii})}},$$

onde $y_i^* = \log\left(\frac{y_i}{1 - y_i}\right)$ e $\hat{\mu}_i^* = \psi(\mu_i\phi) - \psi((1 - \mu_i)\phi)$, em que $\psi(\cdot)$ é a função digama (derivada logarítmica da função Gama). A padronização é dada por $v_i = \psi'(\mu_i\phi) + \psi'((1 - \mu_i)\phi)$ e h_{ii} que diz respeito ao i -ésimo elemento da diagonal da matriz $H = \hat{W}^{1/2} X(X^T \hat{W} X)^{-1} X^T \hat{W}^{1/2}$. \hat{W} é o estimador de máxima verosimilhança de W dada por

$$W = \text{diag}\{w_1, \dots, w_n\}$$

onde $w_i = \phi\{\psi'(\mu_i\phi) + \psi'((1 - \mu_i)\phi)\} \frac{1}{\{g'(\mu_i)\}^2}$. (Consultar Ferrari e Cribari-Neto (2004) e Espinheira et al. (2008a)).

Como se observou no caso do modelo de regressão linear múltipla, também no modelo de regressão beta, podem ser efetuados testes de significância individual dos regressores, com base na estatística de Wald.

É importante a realização de um teste à significância conjunta dos regressores, a fim de averiguar se os coeficientes de regressão são conjuntamente significativos na explicação da variável resposta. O teste de significância conjunta, no âmbito do modelo de regressão beta, é efetuado através da razão de log-verosimilhanças. Segundo o Apêndice B de Ferrari e Cribari-Neto (2004), a estatística da razão de log-verosimilhanças é dada por

$$w_1 = 2\left\{l(\hat{\beta}, \hat{\phi}) - l(\tilde{\beta}, \tilde{\phi})\right\}$$

onde $l(\beta, \phi)$ é a função de log-verosimilhança e $l(\tilde{\beta}^T, \tilde{\phi})^T$ é o estimador de máxima verosimilhança de $(\beta^T, \phi)^T$, sob a hipótese nula, ou seja relativo ao modelo com restrição.

Do mesmo modo que é aplicado o teste da razão de verossimilhanças a todos os coeficientes de regressão do modelo, é possível aplicar este mesmo teste, apenas a um conjunto de regressores, possibilitando a realização de testes aos submodelos.

À semelhança da avaliação do ajustamento dos modelos de regressão linear, os modelos de regressão beta podem também ser avaliados através de uma aproximação das distâncias de Cook, possibilitando assim a identificação de observações *outliers*.

2.3 Método de Seleção de Variáveis

É de grande importância selecionar um subconjunto de variáveis, que, guardando características explicativas e preditivas, seja o mais explicativo da variável resposta. Deste modo, existem métodos que permitem selecionar o melhor subconjunto de variáveis explicativas, entre eles os métodos Backward e o Forward. O método Backward é um método de eliminação, onde o processo de escolha das variáveis baseia-se na eliminação de variáveis do modelo. Por outro lado, o método Forward é um método de seleção, onde, num processo iterativo, se escolhe a melhor variável a ser incluída no modelo. Existe ainda um método de seleção de variáveis, que se baseia em conciliar os dois métodos referidos anteriormente, denominado de Stepwise. Foi esta última metodologia de seleção de variáveis a utilizada neste trabalho, onde são aplicados, simultaneamente, o método Forward e o método Backward.

2.3.1 Seleção Forward

O método de seleção Forward, considera como modelo inicial o modelo de regressão nulo, onde se investiga qual a melhor variável, do conjuntos de variáveis explicativas disponíveis, que deve ser incluída no modelo. A escolha da variável a entrar no modelo é feita com base no teste de razão de verossimilhanças, onde se avalia qual a variável que mais contributo dá ao modelo. Neste método de seleção, é necessário definir como critério de entrada da variável no modelo, um α limite, representado por α_{crit} .

Deste modo, o método de inclusão Forward baseia-se na aplicação dos seguintes passos:

1. Considerar o modelo sem variáveis;
2. Analisar, para todas as variáveis que não estejam no modelo, o respetivo p -value caso essa variável seja adicionada ao modelo;
3. Escolher a variável com p -value mais baixo e que seja inferior a α_{crit} ;
4. Repetir o procedimento até não existirem mais variáveis que verifiquem os requisitos de entrada no modelo.

2.3.2 Eliminação Backward

O método Backward, como referido anteriormente, é um método de eliminação de variáveis. Para tal, o processo é iniciado com todas as variáveis explicativas do modelo e, por etapas, cada uma pode ou não ser eliminada. Tal como no método de inclusão de variáveis Forward, também no método de eliminação Backward, é necessário definir, como critério de saída das variáveis um determinado α_{crit} .

Os passos correspondentes à aplicação do método Backward são:

1. Considerar o modelo com todas as variáveis explicativas;
2. Analisar, para todas as variáveis, o respetivo p -value e remover a variável cujo p -value $< \alpha_{crit}$;
3. Reajustar o modelo e repetir o ponto 2;
4. O método termina quando, para todas as variáveis, se tem que p -value $< \alpha_{crit}$.

2.3.3 Método Stepwise

O método Stepwise caracteriza-se por ser uma combinação dos dois métodos apresentados anteriormente, ou seja, são aplicados, simultaneamente os métodos de inclusão e de exclusão de variáveis. Este método inicia-se com o modelo sem variáveis explicativas, e, à medida que cada variável é adicionada, é feito um teste de exclusão, com o intuito de avaliar qual das variáveis que já se encontrava no modelo se tornou menos significativa.

A aplicação deste método, segue os seguintes passos:

1. Considerar o modelo sem variáveis explicativas;
2. Introduzir, pelo método Forward, a variável mais significativa e reajustar o modelo;
3. Das variáveis já incluídas no modelo, avaliar, pelo método Backward, a menos significativa, eliminando-a do modelo;
4. Reajustar o modelo e repetir os passos 2 e 3;
5. O método termina quando todas as variáveis no modelo têm p -value $< \alpha_{crit}$ e todas as variáveis não incluídas têm p -value $> \alpha_{crit}$.

2.4 Análise da Variação dos Resultados em Função da Dimensão das Amostras

Nos dias de hoje, a facilidade na obtenção de dados, faz com que seja cada vez mais comum, a existência de estudos com base em amostras bastante grandes. Estas amostras de grandes dimensões trazem uma série de vantagens, na medida em que torna-se possível

uma maior quantidade e variedade de estudos sobre os mais diversos assuntos. No entanto, deve-se ter algum cuidado no que diz respeito aos resultados estatísticos obtidos desses estudos, isto porque para amostras muito grandes as conclusões das inferências estatísticas podem ser ineficazes e enganosas. Este problema ocorre, nomeadamente com o valor do p -value. Em grandes amostras, o p -value do modelo tende a tomar valores próximos de zero muito rapidamente. São muitos os estudos com amostras de grande dimensão, que suportam as suas hipóteses apenas em baixos valores do p -value e nos sinais dos coeficientes de regressão, no entanto, como mencionado anteriormente, estas conclusões podem ser enganadoras.

Nos modelos de regressão, a hipótese nula considera, normalmente, os parâmetros de interesse iguais a zero, representando uma ausência de efeito destes parâmetros no modelo.

Segundo Lin et al. (2013), o valor do p -value é baseado em estimadores consistentes, tendo, sob a hipótese $H_0 : \beta = 0$ o seguinte limite

$$\lim_{n \rightarrow \infty} p\text{-value} = \lim_{n \rightarrow \infty} P(|\hat{\beta} - \beta| < \varepsilon) = \begin{cases} 0 & \text{se } \beta \neq 0 \\ 1 & \text{se } \beta = 0 \end{cases}$$

Assim, a não ser que o parâmetro β seja exatamente igual a zero, e nesse caso o p -value seria 1, o p -value tende sempre a tomar o valor zero. Os estimadores consistentes têm erros padrão que diminuem quando a amostra aumenta. Assim, para amostras de grande dimensão, os erros padrão são de tal forma pequenos, que, mesmo distâncias mínimas entre o valor estimado e o valor assumido sob H_0 são consideradas estatisticamente significativas. Por este motivo, deve-se ter especial atenção na magnitude do efeito que as grandes amostras têm, efetuando uma análise do modelo que vá além da rejeição da hipótese nula com base no p -value ou no sinal dos coeficientes de regressão.

São poucos os exemplos de estudos que fazem referência ao problema do p -value em grandes amostras, ao invés disso, dão demasiada importância ao valor do p -value para a interpretação dos resultados.

Alguns autores sugerem alternativas para este problema, que passam pela utilização de intervalos de confiança, pela redução do limiar do nível de significância ou pela importância dada ao conceito de significância prática em alternativa ao conceito de significância estatística.

Em Lin et al. (2013), são apresentadas sugestões que possibilitam uma melhor análise estatística para amostras muito grandes, com o objetivo de minimizar o problema do p -value. Algumas das sugestões apresentadas passam por apresentar a sensibilidade da variável resposta em termos da variação dos regressores, reportar intervalos de confiança

e construir gráficos que relacionem o p -value, os coeficientes estimados e a dimensão da amostra.

Uma das sugestões dos autores, referida no artigo como *effect size*, passa por avaliar qual a amplitude do efeito que uma variação do regressor tem na variável resposta. Ou seja, é avaliada a sensibilidade da variável resposta em termos da variação dos regressores. Este aspeto tem especial interesse nos casos em que são aplicadas transformações à variável resposta e/ou aos regressores.

É ainda sugerida uma análise baseada em **intervalos de confiança**, cujo objetivo passa essencialmente por conseguir avaliar a evolução da amplitude do intervalo de confiança, consoante a dimensão da amostra. Este tipo de análise, possibilita uma maior aplicabilidade dos resultados em estudos posteriores assim como uma maior facilidade em avaliar e tirar conclusões práticas.

A terceira e última sugestão dada pelos autores, passa pela construção de múltiplos **gráficos** de diferentes amostras de dimensões mais reduzidas, de forma a obterem-se níveis de significância mais familiares. Através da construção de múltiplas amostras torna-se possível avaliar a variabilidade dos resultados e, por fim, a criação de amostras com dimensões diferentes, permite evidenciar o problema do p -value.

São quatro os tipos de gráficos sugeridos no artigo em referência:

- Gráfico com intervalo de confiança (*Confidence interval chart*)

Este gráfico exibe os intervalos de confiança dos coeficientes estimados, como função da dimensão da amostra, variando desde uma amostra de tamanho mínimo até uma amostra suficientemente grande. Deste modo, é possível observar a magnitude do coeficiente estimado e a diminuição do erro padrão, à medida que a dimensão da amostra aumenta.

- Gráfico com coeficiente, p -value e dimensão da amostra (*Coefficient/p-value/sample-size chart, CPS chart*)

O gráfico CPS exibe a curva do coeficiente em análise e do respetivo p -value, para amostras de dimensões diferentes. São construídas repetidas amostras de dimensões crescentes, ajusta-se novamente o modelo de regressão e tiram-se os valores do coeficiente estimado e do respetivo p -value associado. Estes valores são posteriormente utilizados para a construção do referido gráfico CPS.

- Gráfico com nível de significância a 1% (*1% significance threshold chart*)

Este gráfico, indica para que tamanho da amostra, o coeficiente de cada variável se torna significativo para um nível de significância de 1%.

2.4. ANÁLISE DA VARIAÇÃO DOS RESULTADOS EM FUNÇÃO DA DIMENSÃO DAS AMOSTRAS

- Gráfico com coeficiente, p -value e dimensão da amostra com simulações de Monte Carlo (*Monte Carlo CPS chart*)

O gráfico CPS de Monte Carlo extrai várias amostras de cada dimensão da amostra, através de repetidas simulações de Monte Carlo, proporcionando assim, informação adicional acerca da distribuição dos coeficientes e dos p -values à medida que a dimensão da amostra aumenta. A implementação deste gráfico é computacionalmente mais intensiva comparativamente aos restantes três gráficos apresentados anteriormente.

Estes gráficos permitem exibir o problema do p -value, mostrando que à medida que o tamanho da amostra aumenta para além de um certo ponto, o p -value cai para valores próximos de zero, onde permanece até à amostra de dimensão máxima.

MODELAÇÃO DOS DADOS

3.1 Informação Disponibilizada pelo Banco de Portugal

É da competência do Banco de Portugal recolher e elaborar estatísticas que possibilitem análises e estudos relativos à economia portuguesa e ao seu sistema financeiro.

Em 2012, foi introduzido um novo requisito de informação estatística que passou a exigir o reporte mensal da informação individual das taxas de juro sobre novas operações de empréstimos a sociedades não financeiras. Deste modo, o Banco exige às instituições que concedem crédito, o reporte mensal relativo a todas as novas operações de crédito. É importante salientar que o referido requisito só se aplica a instituições financeiras que, em cada mês, concedam pelo menos 50 milhões de euros em novas operações de empréstimos a sociedades não financeira.

A implementação deste novo requisito estatístico, permitiu a construção de uma base de micro dados, relativos a cada operação. Esta tipologia de base de dados, ao fornecer informação com elevado detalhe, permite uma maior qualidade dos dados disponíveis, o que por sua vez possibilita uma maior variedade de análises e estudos económicos e financeiros.

Deste modo, é possível cruzar as informações provenientes das bases de micro dados com as restantes bases de dados existentes no Banco, permitindo uma maior e melhor informação disponível. Torna-se possível caracterizar não só o tipo de empréstimo, como também o mutuário e o mutuante do contrato. Assim, tem-se também acesso a informações relevantes sobre a sociedade não financeira que pretende obter crédito e sobre a instituição financeira que lhe concede crédito.

O reporte das estatísticas das taxas de juro bancárias de novos empréstimos é utilizado e analisado pelo Banco de Portugal. A divulgação dos principais indicadores daí resultantes é efetuada através das publicações do Boletim Estatístico, das estatísticas disponibilizadas no *BPstat* assim como das atualizações recorrentes feitas nas Notas de Informação Estatística. O *BPstat* é uma aplicação desenvolvida pelo Banco de Portugal, que permite aceder às suas Bases de Dados de Estatísticas Interativas, possibilitando assim a análise de estatísticas relativas à economia portuguesa.

	Portugal					Área do euro ⁽¹⁾				
	Dez-14	Mai-15	Dez-15	Abr-16	Mai-16	Dez-14	Mai-15	Dez-15	Abr-16	Mai-16
	(taxa acordada anualizada - taa - em percentagem)									
	1	2	3	4	5	6	7	8	9	10
Empréstimos										
1 Sociedades não financeiras	4.09	3.82	2.98	3.26	3.38	2.15	1.99	1.85	1.76	1.72
2 Operações até 1 milhão de euros	4.68	4.27	3.64	3.58	3.52	3.09	2.85	2.60	2.51	2.47
3 Operações acima de 1 milhão de euros	3.48	3.22	2.40	2.78	3.24	1.86	1.65	1.59	1.42	1.38
Particulares:										
4 Habitação	3.00	2.41	2.16	2.02	1.99	2.55	2.20	2.28	2.12	2.04
5 Consumo	8.61	8.21	7.70	7.57	7.53	6.00	6.11	5.74	5.88	5.96
6 Outros fins	4.45	4.47	4.32	4.76	4.50	2.67	2.66	2.47	2.51	2.49
Depósitos										
7 Sociedades não financeiras: com prazo acordado, até 1 ano	0.68	0.41	0.39	0.26	0.25	0.42	0.29	0.21	0.18	0.12
8 Particulares: com prazo acordado, até 1 ano	1.26	0.73	0.53	0.46	0.42	0.98	0.87	0.65	0.60	0.55

Fonte: Banco de Portugal e Banco Central Europeu (com cálculos do Banco de Portugal)

Figura 3.1: Taxas referentes a empréstimos e depósitos de SNF's e particulares

A Figura 3.1 faz referência ao quadro publicado no Boletim Económico e no *BPstat*, onde são analisadas as taxas de juro bancárias de empréstimos e depósitos por parte de sociedades não financeiras e particulares, residentes na área do euro. Uma vez que o presente trabalho incidirá apenas nos empréstimos a sociedades não financeiras, torna-se interessante analisar a Figura 3.2, na medida em que faz referência às taxas de juro relativas a essa categoria de novas operações.

As taxas de juro apresentadas são denominadas de taxas acordadas anualizadas (TAA) e correspondem à média ponderada das taxas de juro individualmente acordadas entre a instituição financeira e a sociedade não financeira relativamente a um dado empréstimo. Na Secção 3.2 será feita uma análise mais aprofundada relativamente a este conceito.

Pela análise gráfica relativa à evolução da taxa acordada anualizada, verifica-se um evidente decréscimo das taxas de juro nos últimos anos. Observa-se que tanto os empréstimos inferiores como superiores a 1 milhão de euros têm taxas mais baixas na área do euro comparativamente a Portugal. Conclui-se ainda que os empréstimos com montantes inferiores a 1 milhão de euros têm taxas de juro mais elevadas, tanto na área do euro como em Portugal, comparativamente a empréstimos com montantes superiores a 1 milhão de euros.

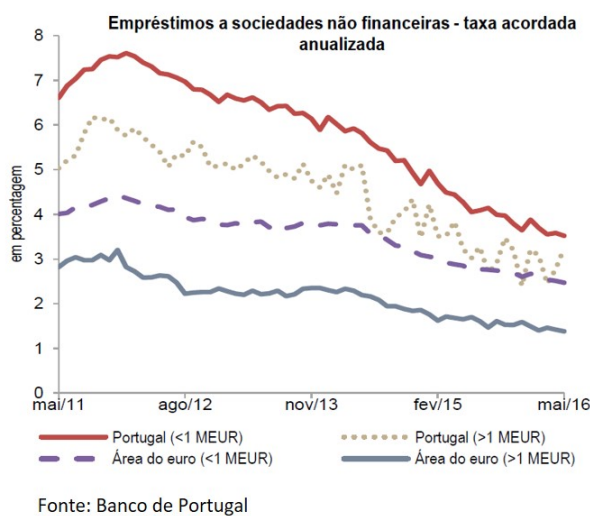


Figura 3.2: Evolução da TAA de empréstimos a sociedades não financeiras

Verifica-se assim, uma distinção na atribuição das taxas de juro consoante o montante do empréstimo. O presente trabalho pretende encontrar outros fatores que sejam também justificativos de diferenças significativas na atribuição das taxas de juro de novas operações de empréstimos.

3.2 Caracterização dos Dados

O objetivo do presente trabalho prende-se com a identificação dos fatores determinantes na fixação das taxas de juro relativas às novas operações de empréstimos, sendo essencial uma análise do comportamento relativo a estes dados.

Tal como referido anteriormente, e de acordo com BdP (2009), as taxas de juro apresentadas são denominadas de taxas acordadas anualizadas (TAA). A TAA, taxa acordada entre a instituição financeira e o cliente, convertida numa base anual, é dada por

$$TAA = \left(1 + \frac{r_{ag}}{n}\right)^n - 1$$

em que

- r_{ag} representa a taxa de juro anual acordada entre a instituição e o cliente, em relação a um dado empréstimo;
- n corresponde ao número de períodos de capitalização de juros durante o ano ($n = 1$ reflete pagamentos anuais, $n = 2$ pagamentos semestrais, $n = 4$ pagamentos trimestrais e $n = 12$ reflete pagamentos mensais).

O requisito estatístico, relativo à informação individual das taxas de juro, exige que seja reportada a média ponderada das taxas pelos respetivos montantes das novas operações, através da seguinte fórmula

$$T = \frac{\sum_j t_j \cdot \text{montante}_j}{\sum_j \text{montante}_j}$$

onde T é a média ponderada das TAA das novas operações, t_j a TAA associada à operação j e montante_j o montante associado à operação j .

O intervalo temporal tido em consideração neste trabalho, diz respeito aos últimos três anos civis, 2013, 2014 e 2015. Neste período, registaram-se mais de 1 milhão de novos empréstimos por parte de cerca de 58 mil empresas, correspondente a um total de aproximadamente 69 mil milhões de euros. A Figura 3.3 indica o número total de novos empréstimos, assim como o número de empresas com novos empréstimos e ainda o volume total dessas operações, para o conjunto global das observações e para cada um dos anos em particular.

Total de observações			
Novas operações:	1 138 629	2013	Novas operações: 391 842
Empresas:	58 783		Empresas: 38 259
Montante:	69 355 M€		Montante: 24 673 M€
		2014	Novas operações: 373 021
			Empresas: 35 602
			Montante: 22 596 M€
		2015	Novas operações: 373 767
			Empresas: 35 758
			Montante: 22 086 M€

Figura 3.3: Caracterização dos dados observados

No decorrer dos três anos, verificou-se uma variação entre os 26 mil e os 38 mil novos empréstimos mensais, por parte de uma média de 13 mil empresas. O volume médio mensal de novos empréstimos, variou entre 1500 e 2500 milhões de euros.

No respeitante às taxas de juro bancárias relativas a novas operações de crédito, estas caracterizam-se por um claro decréscimo, tal como referido anteriormente e como se pode observar na Figura 3.4.

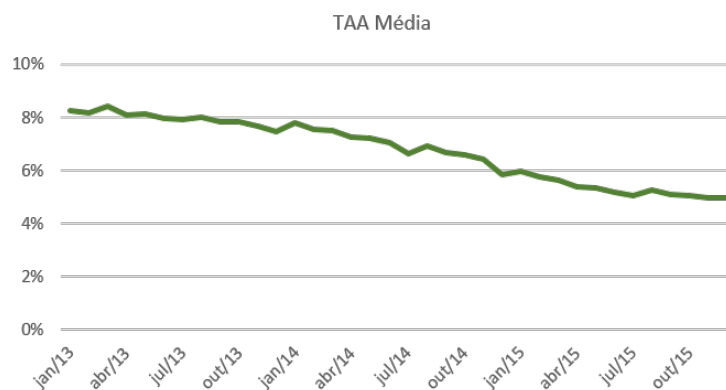


Figura 3.4: Evolução da TAA média

Em Janeiro de 2013 a taxa média das novas operações de empréstimos era de 8.28%, enquanto que em Dezembro de 2015 a mesma taxa média situava-se nos 4.96%, correspondendo assim a uma diminuição de 3.32%. É importante salientar que, num período de apenas três anos, uma diminuição das taxas de juro de 3.3% é considerada uma diminuição bastante acentuada.

O histograma seguinte ilustra a distribuição de frequências da taxa acordada anualizada, TAA, nos três anos em estudo.

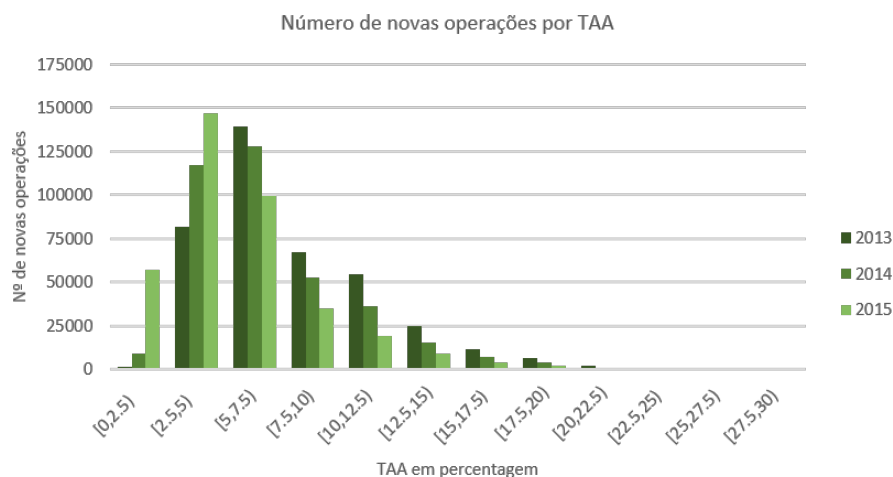


Figura 3.5: Histograma da TAA

Como se pode observar pela Figura 3.5, a distribuição dos dados é assimétrica, verificando-se uma grande concentração de taxas abaixo dos 10%. Observa-se ainda que o número de novos empréstimos com taxas de juro acima dos 20% é muito reduzido. Nos três anos observados, existe um elevado número de novos empréstimos com taxas entre os 2.5% e os 10%, com a particularidade de que no ano de 2015, a concentração de taxas

encontrava-se em valores mais baixos comparativamente aos registados em 2013. Este pormenor realça a queda das taxas de juro ao longo dos três anos.

A base de dados com informação individual das taxas de juro de novas operações de empréstimos a sociedades não financeiras, caracteriza-se por ser uma base de micro dados, com detalhe mensal.

Tal como referido anteriormente, o requisito estatístico referente ao reporte de informação individual das taxas de juro das novas operações de empréstimos, torna possível a caracterização de cada operação de empréstimo através de um conjunto elevado de informação relevante.

A informação disponível nos quadros de reporte das novas operações passa por caracterizar os empréstimos, contendo informação relativa ao montante, à taxa de juro correspondente, ao prazo contratual e ao prazo de fixação inicial da taxa. Identifica-se se o empréstimo tem colateral e/ou garantia e se a operação de empréstimo é totalmente nova ou se é renegociada. Este quadros de reporte indicam ainda a sociedade não financeira através do seu NIPC (Número de Identificação de Pessoa Coletiva), assim como o código da instituição financeira que concede o empréstimo. É através destas duas informações, que se torna possível cruzar dados do empréstimo com informação da empresa e da instituição financeira.

Pretende-se estudar o comportamento das taxas de juro não só consoante as características do empréstimo mas também tendo em atenção as características das empresas. Deste modo, no que diz respeito à informação que se retira da base de micro dados, com informação empréstimo a empréstimo, tem-se os seguintes dados:

- **Data da operação:** Data de início do contrato de empréstimo.
- **Montante:** Valor do empréstimo em milhões de euros.
- **Taxa Acordada Anualizada:** Taxa de juro acordada entre a instituição que concede crédito e a sociedade não financeira, relativamente a cada empréstimo, convertida numa base anual.
- **Prazo contratual:** Maturidade do empréstimo.
- **Prazo de fixação inicial da taxa:** Período definido no início do contrato em que o valor da taxa de juro não é alterado.
- **Empréstimo com colateral:** Empréstimo com ou sem colateral e/ou garantia associada.
- **Tipo de negociação:** São considerados quatro tipos de negociação; operação de empréstimo totalmente nova, renegociação das condições contratuais com o

envolvimento ativo do cliente, renegociação das condições contratuais sem o envolvimento ativo do cliente e nova operação de empréstimo para reestruturação da dívida em que exista um historial de incumprimento associado.

- **Instituição Financeira que concede crédito:** Informação de qual a instituição financeira que concedeu a respetiva operação de empréstimo.

Após o cruzamento de dados, é possível aceder às seguintes características das empresas com novas operações de crédito:

- **Secção da CAE:** Secção da Classificação Portuguesa das Atividades Económicas. As secções da CAE e os respetivos ramos de atividade estão enunciados na tabela 3.1.

Secção CAE	Ramo de Atividade
A	Agricultura, produção animal, caça, floresta e pesca
B	Indústrias extrativas
C	Indústrias transformadoras
D	Eletricidade, gás, vapor, água quente e fria e ar frio
E	Captação, tratamento e distribuição de água e gestão de resíduos
F	Construção
G	Comércio por grosso e a retalho; reparação de veículos
H	Transportes e armazenagem
I	Alojamento, restauração e similares
J	Atividades de informação e de comunicação
K	Atividades financeiras e de seguros
L	Atividades imobiliárias
M	Atividades de consultoria, científicas, técnicas e similares
N	Atividades administrativas e dos serviços de apoio
O	Administração pública, defesa e segurança social
P	Educação
Q	Atividades de saúde humana e apoio social
R	Atividades artísticas, de espetáculos, desportivas e recreativas
S	Outras atividades de serviços
T	Atividades das famílias, atividades de produção para uso próprio
U	Atividades de organismos internacionais e instituições extraterritoriais

Tabela 3.1: Secções CAE

- **Dimensão da empresa:** Micro, pequena, média ou grande empresa. São consideradas micro empresas, todas as sociedades não financeiras com menos de 10 empregados e com volume de negócios e/ou balanço total anual até 2 milhões de euros. São consideradas pequenas empresas, aquelas com menos de 50 empregados, com volume de negócio e/ou balanço total anual até 10 milhões de euros. Empresas médias são as cujo número de trabalhadores é inferior a 250, o volume de negócios

não excede os 50 milhões de euros ou o balanço total anual não excede os 43 milhões de euros. As grandes empresas são todos os restantes casos, ou seja mais de 250 trabalhadores e volume de negócios e balanço total anual superiores aos das médias empresas.

- **Probabilidade de Default / Classe de Rating:** Probabilidade de default e consequente classe de rating em que a sociedade não financeira se situa. A probabilidade de default diz respeito à probabilidade que a empresa tem de entrar em incumprimento, não cumprindo com as suas obrigações. A atribuição das classes é feita de acordo com um modelo teórico desenvolvido pelo Banco de Portugal. A cada classe de rating correspondem as probabilidades de default descritas na figura 3.2.

Classe de Rating	Probabilidade de Default
Classe 1	0.0000% - 0.0039%
Classe 2	0.0039% - 0.0099%
Classe 3	0.0099% - 0.0156%
Classe 4	0.0156% - 0.0249%
Classe 5	0.0249% - 0.0396%
Classe 6	0.0396% - 0.0629%
Classe 7	0.0629% - 0.1000%
Classe 8	0.1000% - 0.1586%
Classe 9	0.1586% - 0.2505%
Classe 10	0.2505% - 0.3937%
Classe 11	0.3937% - 0.6146%
Classe 12	0.6146% - 0.9513%
Classe 13	0.9513% - 1.4570%
Classe 14	1.4570% - 2.2037%
Classe 15	2.2037% - 3.2844%
Classe 16	3.2844% - 4.8136%
Classe 17	4.8136% - 6.9227%
Classe 18	6.9227% - 9.7497%
Classe 19	9.7497% - 99.999%
Classe 20	99.999% - 100.00%

Tabela 3.2: Classes de Rating

- **Crédito vencido no ano anterior:** Empresa com crédito vencido no ano anterior, isto é, com crédito em situação de incumprimento de pagamento.
- **PME Líder:** Pequenas e Médias Empresas classificadas como empresas líderes no mercado.
- **Empresa de elevado crescimento (EEC):** Empresas com mais de 10 trabalhadores cuja variação média anual de volume de negócios é superior a 20% em 3 anos.

- **Exportadora:** Empresa com atividade exportadora, isto é, empresas em que pelo menos 50% do volume de negócios é proveniente de exportações de bens e serviços ou pelo menos 10% do volume de negócios é proveniente de exportações de bens e serviços quando estas são superiores a 150 mil euros.

Ainda assim, com base em ambos os conjuntos de informações apresentados, é possível construir novas variáveis que dão mais alguma informação relevante para o estudo em causa, sendo elas

- **Número de instituições financeiras:** Número total de instituições financeiras com as quais a sociedade não financeira tem operações de crédito, por ano.
- **Número de novas operações:** Número total de novas operações de crédito que a empresa efetua, por ano.
- **Instituições financeiras nacionais e estrangeiras:** Caracterização da instituição financeira como sendo nacional ou estrangeira.
- **Instituições financeiras pertencentes ao Top 5:** Identificação das 5 instituições financeiras com maior volume de negócios.

Antes de se realizar a análise univariada e multivariada, é importante referir que, na elaboração do presente trabalho, não foram tidas em conta as operações reportadas de forma incoerente, assim como as operações com taxas negativas, ocorridas no ano de 2015. Os descobertos bancários também não foram considerados como novas operações de empréstimos na medida em que são operações cujas taxas de juro são muito voláteis e não traduzem as condições normais do mercado. De acordo com a definição dada pelo Banco de Portugal, um descoberto bancário é um contrato pelo qual a instituição de crédito permite a um cliente dispor de fundos que excedem o saldo da respetiva conta à ordem, até um limite definido no respetivo contrato. No intervalo temporal em análise, os descobertos bancários somaram cerca de 60% da totalidade de novas operações de empréstimos concedidos a sociedades não financeiras consideradas pelo Banco de Portugal.

3.3 Análise das Variáveis Explicativas

As variáveis explicativas, tal como referido anteriormente, cobrem um vasto conjunto de características relativas aos empréstimos e também às empresas que efetuem essas operações. De seguida, efetuar-se-á uma abordagem descritiva das variáveis explicativas consideradas, identificando a sua natureza e, em certos casos, a codificação utilizada. Será apresentado um conjunto de gráficos que caracterizarão as variáveis explicativas consideradas, relativamente aos empréstimos e às empresas.

3.3.1 Características do Empréstimo

Nesta secção são apresentadas as variáveis explicativas que caracterizam as operações de empréstimo, características essas que passam pelo montante, pela maturidade, pela entidade que cede o empréstimo, assim como pelo tipo de negociação da operação. As variáveis apresentadas são tanto de carácter quantitativo como de carácter qualitativo e os dados relativos a estas características são reportados operação a operação.

As variáveis explicativas respeitantes às características dos empréstimos, assim como as respetivas descrições, encontram-se resumidas na Tabela 3.3.

Variável	Descrição
Montante	Variável contínua que indica o logaritmo do montante do empréstimo, em milhões de euros
Prazo Contratual	Variável categórica que indica a maturidade do empréstimo 0: Até 3 meses 1: De 3 a 12 meses 2: De 1 a 5 anos 3: Mais de 5 anos
Prazo de Fixação Inicial	Variável categórica que indica o prazo de fixação inicial da taxa de juro relativa ao empréstimo 0: Até 3 meses 1: De 3 a 12 meses 2: De 1 a 5 anos 3: Mais de 5 anos
Tipo de Negociação	Variável categórica que indica o tipo de negociação da operação 0: Totalmente nova 1: Renegociada com o envolvimento ativo do cliente 2: Renegociada sem o envolvimento ativo do cliente 3: Para reestruturação da dívida, em que exista historial de incumprimento associado
Colateral	Variável binária que indica se o empréstimo tem colateral 0: Empréstimo sem colateral e/ou garantia 1: Empréstimo com colateral e/ou garantia
Top 5	Variável categórica que indica a instituição financeira que concede o empréstimo 0: Se a instituição financeira não faz parte das 5 IF's com maior volume de negócios 1, 2, 3, 4 ou 5: Se a instituição financeira faz parte das 5 maiores a operar em Portugal
IF Nacional vs Estrangeira	Variável binária que indica se a instituição financeira que concede o empréstimo é nacional ou estrangeira 0: Instituição financeira nacional 1: Instituição financeira estrangeira

Tabela 3.3: Variáveis explicativas relativas às características do empréstimo

3.3.1.1 Montante

Uma das características mais relevantes no que diz respeito a um empréstimo, é o seu montante associado. Por este motivo, é de grande interesse efetuar uma análise descritiva inicial dos dados relativamente a esta variável, com o intuito de aferir qual o seu comportamento.

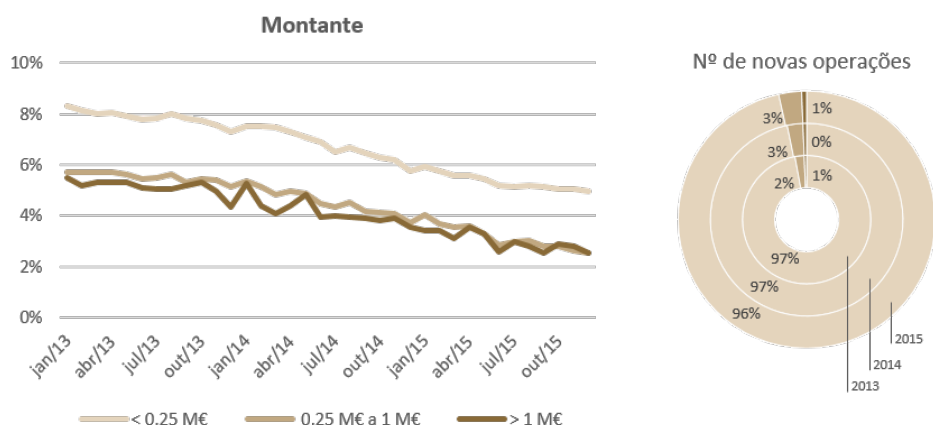


Figura 3.6: Evolução das taxas consoante o montante associado ao empréstimo

Tal como visto anteriormente, a tendência geral das taxas de juro no período de tempo em análise foi de um explícito decréscimo, facto que se pode observar também no gráfico da Figura 3.6. Para a análise descritiva desta variável, optou por se distinguir os empréstimos em três intervalos consoante o montante associado, tal como é feito no reporte de informação estatística. Deste modo, os empréstimos podem ser inferiores a 250 mil euros, de 250 mil a 1 milhão de euros e superiores a 1 milhão de euros. É possível observar que, relativamente aos dados observados, os empréstimos abaixo dos 250 mil euros são os que têm taxas mais elevadas, verificando-se que quanto maior o montante de um empréstimo, menor é a taxa de juro associada. Este fenómeno poderá estar relacionado com o facto de empréstimos de maior montante exigirem uma análise mais rigorosa e pormenorizada, sendo necessário preencher um maior conjunto de requisitos, de forma a ser concedido o empréstimo. O gráfico circular evidencia a composição dos dados, em cada um dos anos, distinguindo os três intervalos de montantes. Verifica-se uma clara predominância de empréstimos com montantes inferiores a 250 mil euros, existindo, anualmente, apenas cerca de 3% de empréstimos com montantes superiores a 250 mil euros. Os empréstimos superiores a 1 milhão de euros, no ano de 2015, corresponderam apenas a 1% do total de novos empréstimos.

3.3.1.2 Prazo Contratual

A maturidade, ou prazo contratual, corresponde ao prazo de vencimento do empréstimo e é também uma variável muito importante na atribuição das taxas de juro. A Figura

3.7 ilustra a evolução das taxas consoante os diferentes prazos contratuais, assim como também caracteriza as novas operações de acordo com o prazo contratual e os respetivos montantes associados.

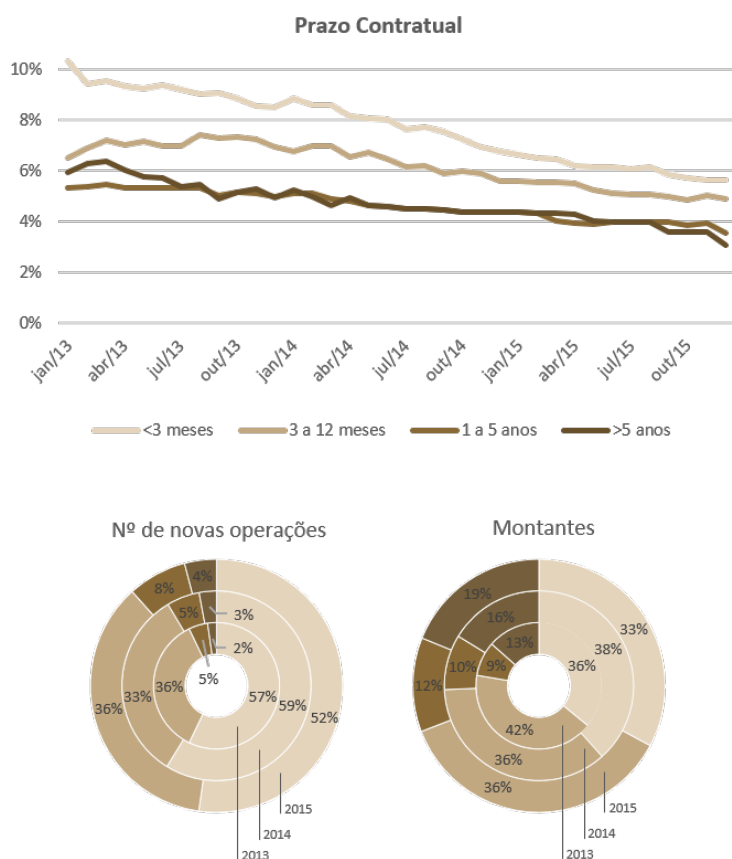
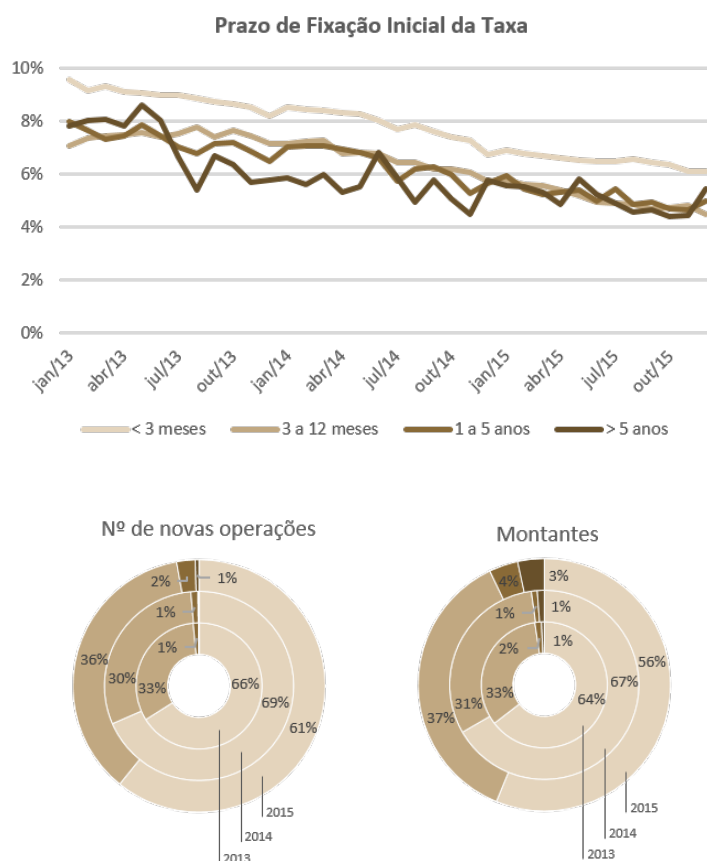


Figura 3.7: Evolução das taxas consoante o prazo contratual associado ao empréstimo

Pode observar-se que a maioria dos empréstimos são de curto prazo, com maturidade inferior a 3 meses, existindo ainda uma percentagem considerável de empréstimos superiores a 3 meses e inferiores a 1 ano. A percentagem de operações com prazos contratuais superiores a 1 ano é baixa, sendo ainda mais reduzida quando se consideram empréstimos a mais de 5 anos. No entanto, observa-se que apesar de serem poucos, empréstimos com prazos mais alargados têm montantes mais elevados.

3.3.1.3 Prazo de Fixação Inicial da Taxa

O prazo de fixação inicial da taxa corresponde ao prazo definido no início do contrato, durante o qual a taxa de juro não pode ser alterada. A divisão do prazo de fixação da taxa em intervalos, foi feita de forma análoga à divisão do prazo contratual, resultando os mesmos intervalos de tempo, até 3 meses, de 3 meses a 1 ano, de 1 a 5 anos e a mais de 5 anos. É importante referir que o prazo de fixação inicial da taxa é sempre igual ou inferior ao prazo contratual. O gráfico relativo à evolução das taxas de juro, Figura



© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

1000

1000

só a partir de 2015 se começaram a efetuar, com maior regularidade, empréstimos com essa característica. Pela análise do gráfico da Figura 3.9 observa-se que as taxas relativas a empréstimos resultantes de renegociações automáticas são bastante mais elevadas que as restantes, isto porque o cliente não consegue intervir na negociação da taxa. Por oposição, os empréstimos que resultam de uma renegociação com o envolvimento ativo do cliente, são os que têm taxas de juro mais baixas. Relativamente ao número de novas operações, verifica-se que a grande maioria das operações diz respeito a empréstimos totalmente novos, assim como os respetivos montantes.

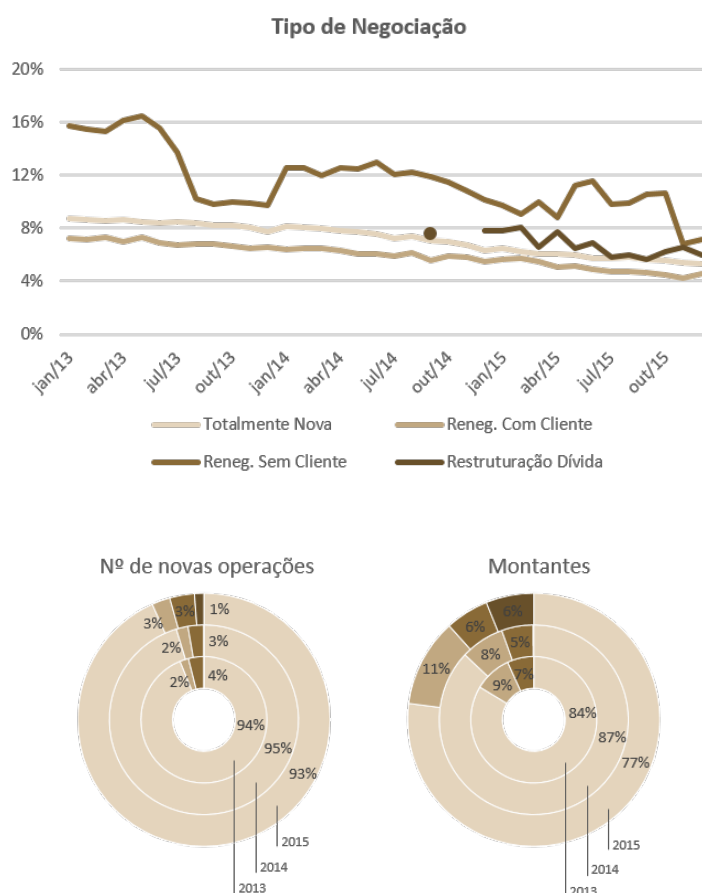


Figura 3.9: Evolução das taxas consoante o tipo de negociação do empréstimo

3.3.1.5 Empréstimo com Colateral

A variável explicativa relativa ao colateral, distingue os empréstimos em que existe colateral e/ou garantia, dos empréstimos onde esta característica não se verifica. A Figura 3.10 ilustra a evolução das taxas de juro para empréstimo com e sem colateral. Em ambos os casos é verificada a tendência de decréscimo das taxas de juro, não sendo evidente a existência de diferenças significativas entre os dois cenários. Relativamente à composição das observações, verificou-se que, anualmente, cerca de 60% das novas operações dizem

respeito a empréstimos sem colateral. Os empréstimos com colateral ou garantia, apesar de serem em menor número, têm montantes superiores. É possível a interpretação de que em empréstimos com montantes mais elevados seja exigido colateral com maior frequência comparativamente a empréstimos de menor montante.

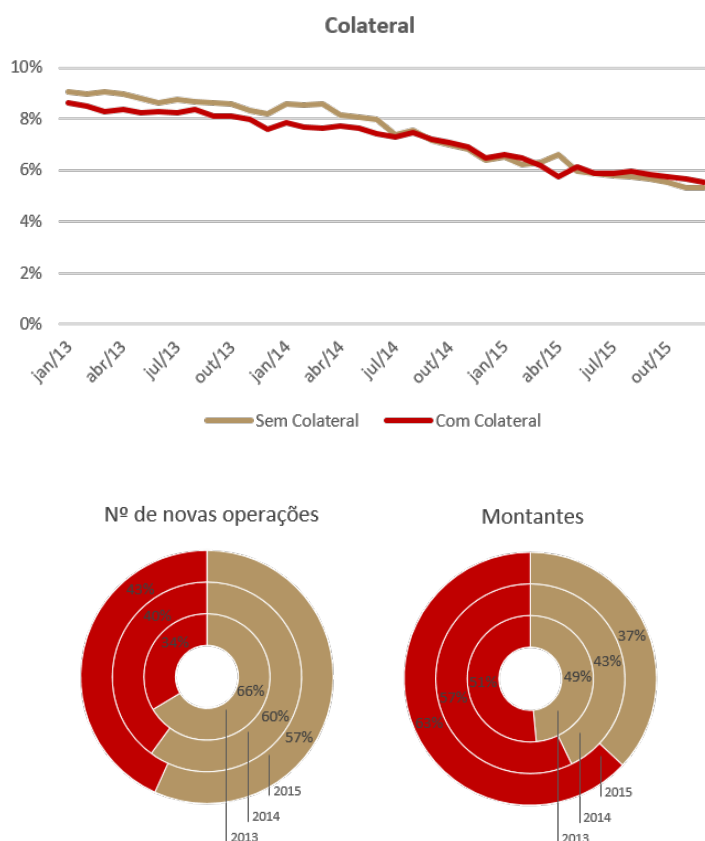


Figura 3.10: Evolução das taxas com a distinção de empréstimos com e sem colateral

3.3.1.6 Bancos pertencentes ao Top 5

O reporte de informação individual das taxas de juro de empréstimos bancários é obrigatória por parte das instituições financeiras, tornando-se interessante analisar o comportamento das taxas de juro consoante o banco que concede crédito. Foram considerados os 5 maiores bancos a operar em Portugal (denominados no presente trabalho de "Bancos Top 5") com o intuito de poder avaliar e comparar as taxas de juro de novas operações de empréstimos, em cada um desses cinco bancos. Por uma questão de confidencialidade, os bancos pertencentes ao Top 5 são os Bancos A, B, C, D e E. Através da Figura 3.11, observam-se algumas diferenças no que toca às taxas de juro aplicadas por cada um dos bancos pertencentes ao Top 5. Verifica-se que, em média, os bancos que não pertencem ao Top 5 têm taxas elevadas, quase tão altas como um dos 5 maiores bancos, o Banco D. Os restantes 4 bancos, aplicam, em média, taxas de juro mais baixas, sendo o Banco B e o Banco C os que aplicam as taxas mais reduzidas.

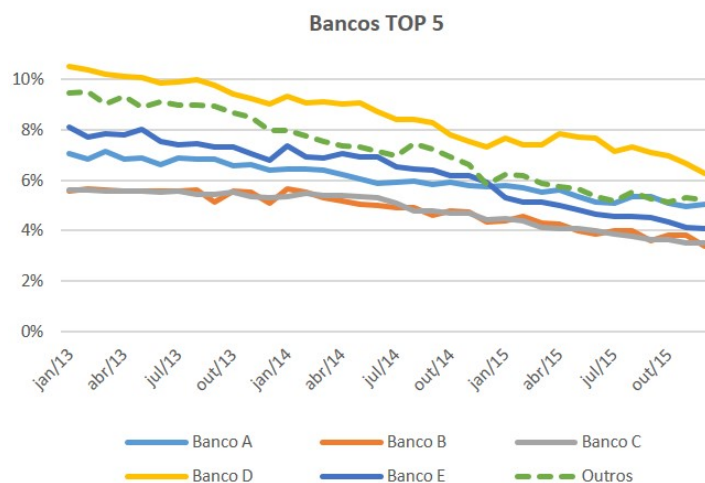


Figura 3.11: Evolução das taxas consoante os bancos pertencentes ao Top 5

3.3.1.7 IF Nacional vs. Estrangeira

Uma distinção interessante a fazer relativamente à instituição financeira que concede crédito está relacionada com a sua nacionalidade. Deste modo, e dada a importância da nacionalidade da instituição financeira, procedeu-se à comparação de instituições financeiras nacionais e estrangeiras. Através da análise do gráfico da Figura 3.12, constata-se que, em média, as instituições financeiras nacionais fixam taxas mais elevadas que as fixadas pelas instituições financeiras estrangeiras. É de notar que, para os dados observados, a grande maioria das novas operações de empréstimos bancários foram realizadas em instituições financeiras nacionais, correspondendo uma menor percentagem às operações realizadas em instituições estrangeiras.



Figura 3.12: Evolução das taxas com a distinção de IF's nacionais e estrangeiras

3.3.2 Características da Empresa

Relativamente às empresas que realizam novas operações de empréstimo, as variáveis explicativas apresentadas passam pela caracterização da empresa, do tipo de atividade ou da dimensão que têm.

Variável	Descrição
Dimensão da empresa	Variável contínua que indica o logaritmo do total de ativo que a empresa possui (um dos fatores que caracteriza e identifica a dimensão da empresa como micro, pequena, média ou grande)
Setor de Atividade	Variável categórica que indica o ramo de atividade da empresa A: Agricultura BC: Indústrias Extrativas e Transformadoras DE: Eletricidade, Gás e Água FL: Construção e Atividades Imobiliárias H: Transportes e Armazenamento I: Alojamento e Restauração JKMN: Atividades de Informação, Financeiras, Científicas e Administrativas OPQ: Administração Pública, Segurança Social, Educação e Saúde RSTU: Outras Atividades e Serviços
Classe de Rating	Variável categórica que indica a classe de rating da empresa 1: Empresa classificada entre a classe 1 e a classe 5 2: Empresa classificada entre a classe 6 e a classe 10 3: Empresa classificada entre a classe 11 e a classe 15 4: Empresa classificada entre a classe 16 e a classe 20
Crédito Vencido	Variável binária que indica se a empresa teve crédito vencido 0: A empresa não teve crédito vencido 1: A empresa teve crédito vencido
PME Líder	Variável binária que indica se a empresa é PME Líder 0: A empresa não é PME Líder 1: A empresa é PME Líder
Exportadora	Variável binária que indica se a empresa é exportadora 0: A empresa não é exportadora 1: A empresa é exportadora
EEC	Variável binária que indica se a empresa tem a classificação de empresa de elevado crescimento (EEC) 0: A empresa não é EEC 1: A empresa é EEC
Número de IF's	Variável discreta que indica o número de instituições financeiras com as quais a empresa tem novos empréstimos no corrente ano
Número de Novas Operações	Variável contínua relativa ao logaritmo do número total de novas operações de empréstimo que a empresa tem no corrente ano

Tabela 3.4: Variáveis explicativas relativas às características da empresa

Tal como se verificou nas características dos empréstimos, também aqui, são apresentadas variáveis quantitativas e qualitativas. As características relativas às empresa são anuais, na medida em que, grande parte das informações são reportadas anualmente pela IES (Informação Anual Simplificada) e, portanto, só se verificam alterações de ano para ano. Estas variáveis explicativas são descritas na Tabela 3.4.

3.3.2.1 Dimensão

Uma das principais características de todas as empresas é a sua dimensão. Quanto a este atributo, as empresas podem ser classificadas como micro, pequenas, médias ou grandes.

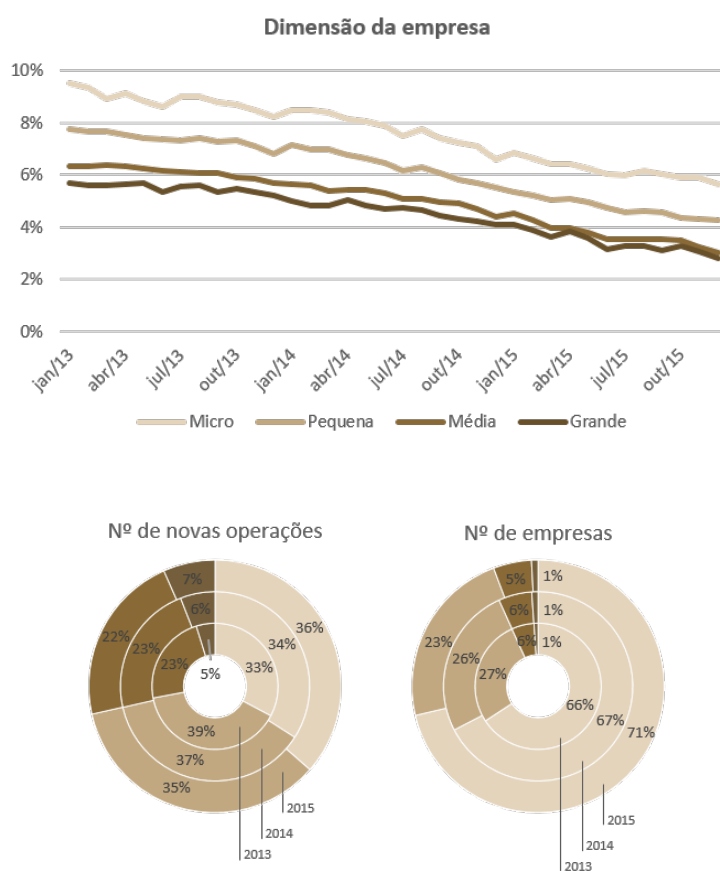


Figura 3.13: Evolução das taxas consoante a dimensão da empresa

O gráfico da Figura 3.13, permite observar a distinção das taxas consoante a dimensão da empresa, verificando que quanto maior é a dimensão, mais baixa é a taxa de juro atribuída. Desta forma, as micro empresas têm, em média, taxas mais elevadas em oposição às grandes empresas que têm as taxas mais reduzidas. Analisando o gráfico circular correspondente ao número de empresas, observa-se que cerca de 70% das empresas com novos empréstimos são micro empresas e que apenas 1% corresponde a grandes empresas. No entanto, o número de novas operações por parte de cada dimensão de empresa apresenta algumas diferenças. As micro e pequenas empresas somam 70% do

total de novas operações. Por outro lado, as grande empresas, que correspondem apenas a 1% do total de empresas analisadas, têm cerca de 6% no total do número de novas operações.

3.3.2.2 Setor de Atividade

O tipo de atividade a que a empresa se dedica é outra característica que pode ter influência na atribuição das taxas de juro de novas operações de empréstimos. Os setores de atividade definidos pelas secções da CAE são 21, mas, no presente trabalho, foram efetuadas algumas agregações de secções com base em dois aspetos, taxas com comportamentos semelhantes ou setores com atividades semelhantes. Um exemplo desta agregação são o setor B, Indústrias Extrativas, e o setor C, Indústrias Transformadoras, que neste trabalho foram agregados, resultando o setor de atividade BC, relativo a Indústrias Extrativas e Transformadoras. Foram também agregadas as secções D e E, as secções F e L, as secções J, K, M e N, as secções O, P e Q e, finalmente as secções R, S, T e U.

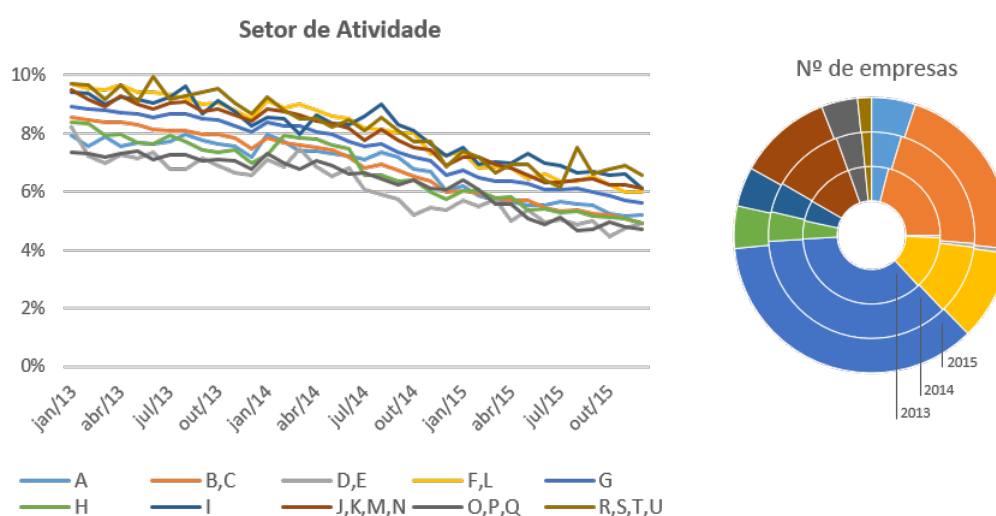


Figura 3.14: Evolução das taxas consoante o setor de atividade da empresa

O gráfico circular da Figura 3.14, respeitante à distribuição das empresas pelos setores de atividade, permite observar que existem mais empresas com novas operações de empréstimos no setor BC, relativo às Indústrias Extrativas e Transformadoras, e no setor G, relativo ao Comércio. No que diz respeito à variação das taxas consoante o setor de atividade, é possível observar algumas diferenças, nomeadamente que os setores DE (Eletricidade, Gás, Água, Saneamento e Gestão de resíduos) e OPQ (Administração Pública, Segurança Social, Educação e Saúde) têm taxas relativamente baixas, quando comparadas com as dos setores FL (Construção e Atividades Imobiliárias) e I (Alojamento e Restauração).

3.3.2.3 Rating

Uma informação bastante utilizada para caracterizar as empresas é a sua classe de rating, que indica o risco que a empresa tem em não cumprir com as suas obrigações. Para o presente trabalho, as 20 classes de rating consideradas pelo Banco de Portugal, foram agrupadas em 4 grupos, do menor para o maior risco.

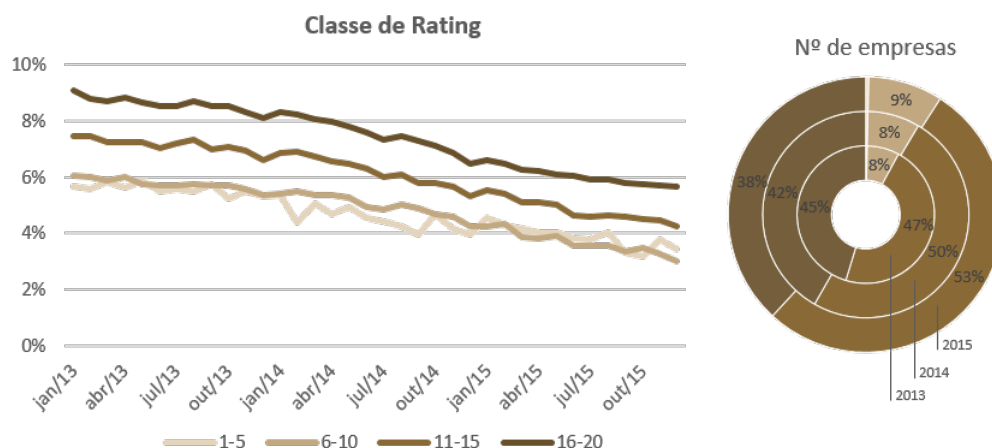


Figura 3.15: Evolução das taxas consoante a classe de rating da empresa

A Figura 3.15 evidencia a diferença das taxas atribuídas consoante a classe de rating em que a empresa se encontra. As taxas mais elevadas são atribuídas a empresas com rating entre a classe 16 e a 20, ou seja, empresas com elevado risco. Por outro lado, as taxas mais reduzidas são atribuídas, como esperado, a empresas com risco mais baixo, ou seja com uma classificação de rating entre a classe 1 e 5.

A percentagem de empresas com classificação de rating entre as classes 1 e 5 é quase nula, existindo escassas empresas com um baixo risco de incumprimento. Verificou-se que cerca de 9% das empresas tem um risco moderadamente baixo, entre a classe 6 e a classe 10, e 50% tem um risco moderadamente alto de incumprimento, entre a classe 11 e a classe 15. No que diz respeito a um risco de incumprimento bastante elevado, da classe 16 à classe 20, foram cerca de 40% as empresas observadas.

3.3.2.4 Crédito Vencido

De acordo com a informação disponível, o facto de a empresa ter tido, ou não, crédito vencido no ano anterior ao empréstimo, tem um impacto na atribuição da taxa de juro, tal como se pode observar no gráfico da Figura 3.16. Caso a empresa tenha tido crédito vencido, a taxa atribuída é mais elevada, comparativamente ao caso em que a empresa não tenha tido crédito vencido. Uma empresa tem crédito vencido se não cumpriu com as suas obrigações, o que, aos olhos da instituição financeira, é considerado como um comportamento de risco, justificando assim a atribuição de taxas mais elevadas.

Com base nos dados analisados, pode ainda observar-se que a grande maioria das empresas analisadas não teve crédito vencido no ano anterior ao da realização do empréstimo.

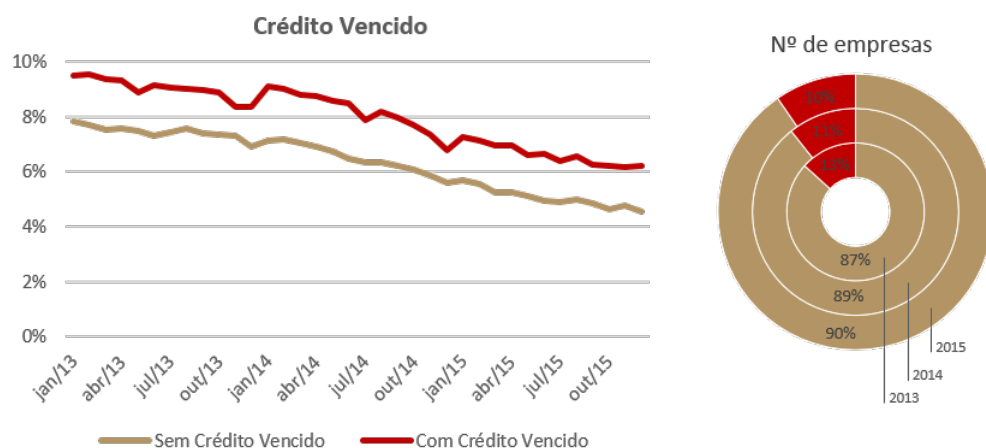


Figura 3.16: Evolução das taxas com a distinção da empresa ter ou não crédito vencido

3.3.2.5 PME Líder

A variável PME líder, Figura 3.17, caracteriza a empresa no caso de ser, ou não, considerada uma pequena ou média empresa líder. De acordo com o gráfico que traduz a evolução das taxas, verifica-se que são atribuídas taxas de juro mais baixas às empresas consideradas PME líderes e taxas mais elevadas às empresas que não possuem esta característica. Observa-se ainda que, nos três anos em estudo, a percentagem de empresas consideradas como PME líderes foi bastante baixa, cerca de 12% por ano.

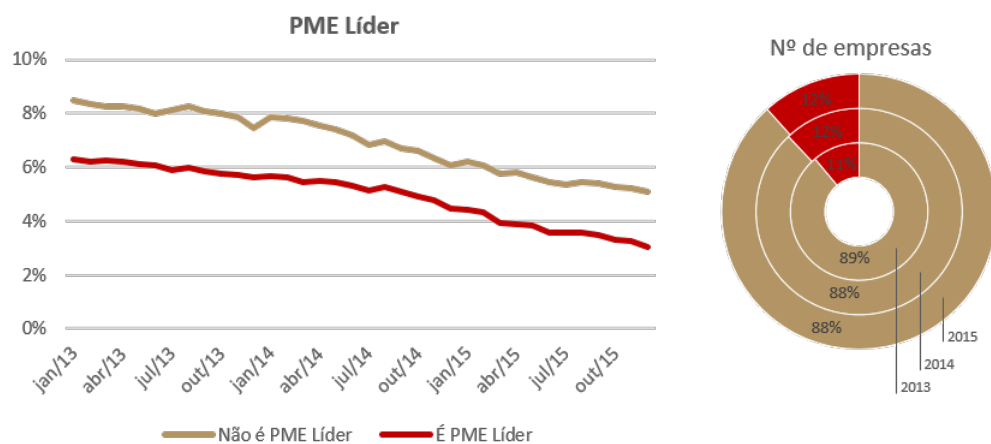


Figura 3.17: Evolução das taxas consoante a característica PME líder

3.3.2.6 Empresa Exportadora

A atividade exportadora é, também ele, um fator importante na caracterização de uma empresa e, tal como se observa na Figura 3.18, parece ser um fator distintivo das taxas de juro em novas operações de empréstimos. De acordo com o gráfico, verifica-se que as empresas com atividade exportadora têm, em média, taxas mais baixas comparativamente às empresas sem atividade exportadora. Através do gráfico circular, observa-se uma baixa percentagem de empresas exportadoras no conjunto total dos dados, evidenciando que a maioria das empresas analisadas não tem atividade exportadora, cerca de 95% por ano.



Figura 3.18: Evolução das taxas consoante a característica relativa às exportações

3.3.2.7 Empresa de Elevado Crescimento

A variável EEC corresponde a uma característica anual que indica se a empresa tem a classificação de empresa de elevado crescimento.

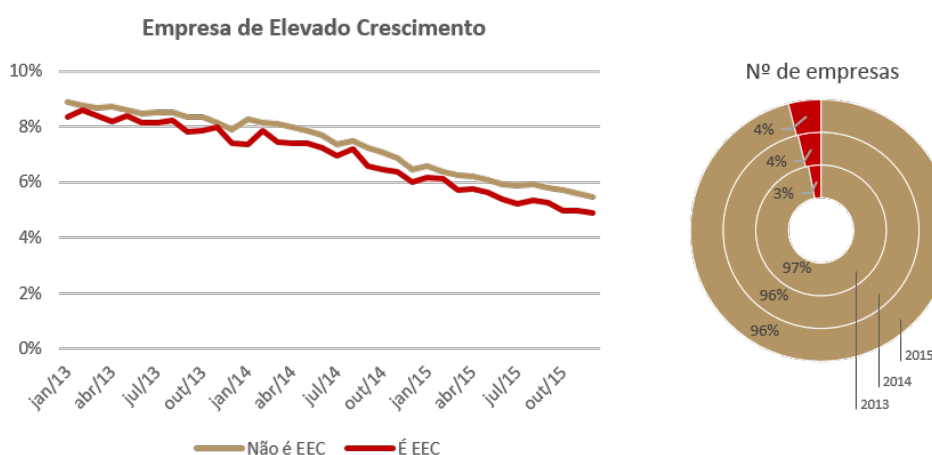


Figura 3.19: Evolução das taxas consoante a característica EEC

A Figura 3.19 ilustra a evolução das taxas consoante a empresa tenha, ou não, elevado crescimento. Verifica-se que para empresas com elevado crescimento as taxas de juro atribuídas às novas operações de empréstimos bancários são ligeiramente mais baixas comparativamente às empresas sem elevado crescimento.

3.3.2.8 Número de Instituições Financeiras

A variável relativa ao número de instituições financeiras corresponde ao número de IF's com as quais a empresa tem novas operações de empréstimos bancários.

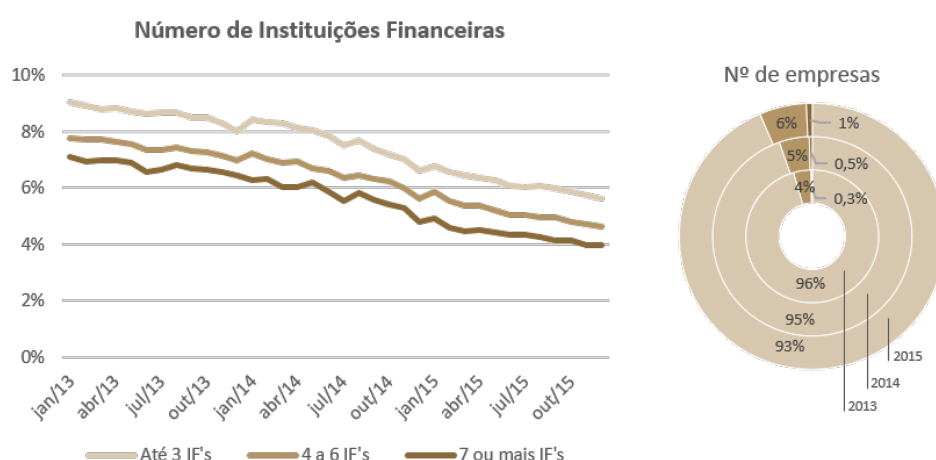


Figura 3.20: Evolução das taxas consoante o número de instituições financeiras

Segundo a análise gráfica, Figura 3.20, há indicação de que com quantas mais instituições financeiras a empresa tenha empréstimos, mais baixas são as taxas de juro conseguidas. Constata-se que mais de 90% das empresas com novas operações de empréstimos têm, no máximo, empréstimos em três instituições financeiras diferentes, por ano.

3.3.2.9 Número de Novas Operações

A última característica apresentada é o número de novas operações que a empresa tem no corrente ano. Mais de metade das empresas analisadas, obteve, em cada um dos três anos, até 10 novas operações de empréstimo, cerca de 30% das empresas teve entre 10 e 100 novos empréstimos e apenas cerca de 5% obteve mais de 100 empréstimos.

Relativamente à distinção das taxas de juro atribuídas, não é possível verificar uma distinção clara entre empresas com diferentes números de novas operações, tal como se observa na Figura 3.21.

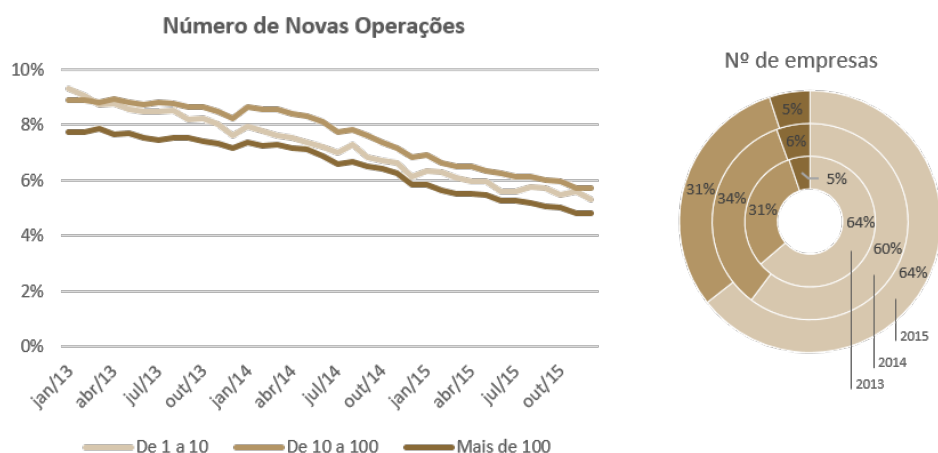


Figura 3.21: Evolução das taxas consoante o número de novas operações

3.4 Análise Univariada

A análise univariada permite efetuar uma primeira análise dos dados, avaliando a informação disponível. Deste modo, recorre-se a esta análise com o intuito de identificar e quantificar o efeito individual de cada variável na explicação do comportamento das taxas de juro bancárias de novos empréstimos.

Tal como referenciado anteriormente, são dois os modelos de regressão em estudo, o modelo de regressão linear múltipla com transformação de Box-Cox e o modelo de regressão beta. Por este motivo, apresentar-se-á de seguida a análise univariada relativa a estes dois modelos.

Pela Tabela 3.5 conclui-se que todas as variáveis são individualmente significativas, tanto no modelo de regressão linear múltipla com transformação de Box-Cox como no modelo de regressão beta.

Variável	Modelo de Regressão Linear		Modelo de Regressão Beta	
	Coeficiente	Signif.	Coeficiente	Signif.
<i>intercept</i>	-1,7527	***	-0,8531	***
ano2014	-0,0641	***	-0,0495	***
ano2015	-0,1748	***	-0,1349	***
<i>intercept</i>	-1,9669	***	-1,1122	***
log(Montante)	-0,0612	***	-0,0510	***
<i>intercept</i>	-1,6257	***	-0,8159	***
contr1	-0,1449	***	-0,1208	***
contr2	-0,2654	***	-0,2219	***
contr3	-0,2781	***	-0,2309	***

Variável	Modelo de Regressão Linear		Modelo de Regressão Beta	
	Coeficiente	Signif.	Coeficiente	Signif.
<i>intercept</i>	-1,7249	***	-0,8821	***
fix1	-0,1043	***	-0,0819	***
fix2	-0,1390	***	-0,1106	***
fix3	-0,1629	***	-0,1268	***
<i>intercept</i>	-1,7702	***	-0,9178	***
OpReneg1	-0,0757	***	-0,0552	***
OpReneg2	0,1247	***	0,1031	***
OpReneg3	-0,0036		0,0013	
<i>intercept</i>	-1,6817	***	-0,8599	***
colateral1	-0,0967	***	-0,0769	***
<i>intercept</i>	-1,7666	***	-0,9052	***
top5_1	-0,0482	***	-0,0327	***
top5_2	-0,1794	***	-0,1334	***
top5_3	-0,0889	***	-0,0731	***
top5_4	0,0932	***	0,0767	***
top5_5	0,0393	***	0,0329	***
<i>intercept</i>	-1,7958	***	-0,8975	***
IFnac1	-0,1085	***	-0,0874	***
<i>intercept</i>	-1,5059	***	-0,6862	***
log(TA)	-0,0217	***	-0,0168	***
<i>intercept</i>	-1,7688	***	-0,9024	***
setorA	-0,0692	***	-0,0532	***
setorBC	-0,0024		-0,0028	
setorDE	-0,0754	**	-0,0588	**
setorFL	0,008		0,0054	
setorH	-0,0275	**	-0,0233	**
setorI	-0,0099		-0,0075	
setorJKMN	-0,0191	**	-0,0147	**
setorOPQ	-0,0824	***	-0,0637	***
setorRSTU	0,0009		0,0030	
<i>intercept</i>	-1,7280	***	-0,8589	***
rating1	-0,2021	***	-0,1569	***
rating2	-0,1985	***	-0,1527	***
rating3	-0,1028	***	-0,0815	***
<i>intercept</i>	-1,7927	***	-0,9221	***
CV1	0,1184	***	0,0961	***
<i>intercept</i>	-1,7845	***	-0,9023	***
PMElider1	-0,0990	***	-0,0755	***

	Modelo de Regressão Linear		Modelo de Regressão Beta	
Variável	Coeficiente	Signif.	Coeficiente	Signif.
<i>intercept</i>	-1,7780	***	-0,9097	***
EEC	-0,0348	***	-0,0264	**
<i>intercept</i>	-1,7765	***	-0,9086	***
Export1	-0,0644	***	-0,0488	***
<i>intercept</i>	-1,8019	***	-0,9897	***
log(NNO)	0,0718	***	0,0617	***
<i>intercept</i>	-1,8444	***	-0,9793	***
NumIF	0,0527	***	0,0434	***

Tabela 3.5: Análise univariada dos dois modelos em estudo

A fim de averiguar a dependência de variáveis, efetuou-se uma análise de correlações entre as variáveis respeitantes às características dos empréstimos, assim como ao conjunto de variáveis correspondentes às características das empresas.

Verificou-se a existência de correlação entre o prazo contratual do empréstimo e o prazo de fixação inicial da taxa de juro. O prazo de fixação inicial da taxa é sempre igual ou inferior ao prazo contratual, e, são bastantes os casos em que ambos os prazos têm a mesma duração, derivado de existirem bastantes operações de curto prazo. Por este motivo, parece adequado considerar apenas uma das variáveis. Optou-se por considerar a variável relativa ao prazo contratual, na medida em que permite uma melhor caracterização do empréstimo, comparativamente ao prazo de fixação inicial da taxa. As restantes variáveis não evidenciam existência de correlação.

3.5 Análise Multivariada

A utilização de métodos estatísticos que estudem as relações existentes entre grupos de variáveis, e não apenas entre variáveis individuais é de elevado interesse. A abordagem multivariada é bastante importante quando se pretende explicar um fenómeno através de dois ou mais fatores explicativos. Este tipo de análise avalia o efeito da interação de todas as variáveis explicativas na interpretação da variável resposta em estudo.

Como referido anteriormente, pretende-se explicar os fatores determinantes na atribuição das taxas de juro de novas operações de empréstimos bancários a sociedades não financeiras, sendo esta a variável resposta em estudo.

A modelação da variável resposta, tal como se constata pela análise gráfica da Figura 3.22, pode ser realizada utilizando duas distribuições distintas. As distribuições teóricas escolhidas para esta análise multivariada foram as distribuições Normal e Beta.

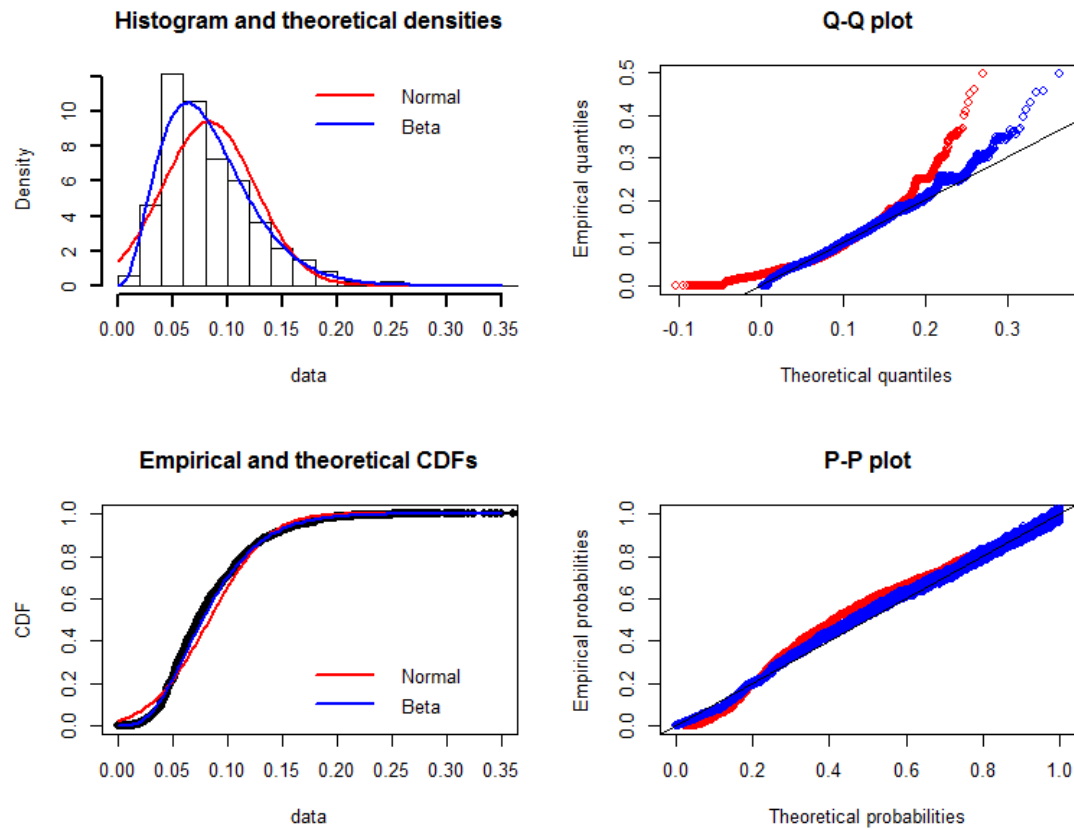


Figura 3.22: Análise do ajustamento das distribuições teóricas à distribuição empírica

A Figura 3.22 possibilita uma análise relativa à qualidade de ajustamento das distribuições teóricas aos dados em análise, Delignette-Muller, Dutang et al. (2015). O primeiro gráfico diz respeito às densidades das duas distribuições teóricas ajustadas ao histograma da distribuição empírica. O gráfico relativo às CDF's, *cumulative distribution function*, exibe, tal como o nome indica, a função de distribuição acumulada dos dados e das duas distribuições ajustadas. O gráfico Q-Q faz a comparação dos quantis teóricos com os quantis empíricos, enquanto que o gráfico P-P compara as probabilidades das distribuições teóricas com as probabilidades da distribuição empírica.

Os gráficos apresentados mostram que a distribuição Beta tem um ajuste razoável e, por esse motivo, considerou-se este modelo para analisar os fatores determinantes no comportamento das taxas de juro. Pelo gráfico Q-Q, verifica-se uma certa falta de ajustamento nas caudas da distribuição Normal, que pode ser justificado por uma falta de simetria das observações. Uma possível solução para a falta de ajustamento da distribuição Normal, passa pela utilização de transformações da variável resposta, como por exemplo a transformação de Box-Cox, apresentada na Subsecção 3.5.2. Voltaremos a este assunto mais à frente.

Relativamente ao conjunto de todas as variáveis explicativas apresentadas na Secção 3.3, verificou-se, como referido anteriormente, a existência de correlação entre o prazo contratual do empréstimo e o prazo de fixação inicial da taxa. Por este motivo, decidi não se incluir a variável relativa ao prazo de fixação, uma vez que se considera o prazo contratual bastante mais importante na caracterização dos empréstimos.

A identificação do ano em que se realiza o empréstimo é um fator importante na explicação das taxas de juro. Por esse motivo, incluiu-se no modelo uma variável categórica, denominada de Ano, correspondendo ao ano de 2013, 2014 ou 2015.

Ao iniciar a construção dos modelos, é importante seleccionar as variáveis explicativas que devem ser incluídas. O método de seleção utilizado no presente trabalho foi o método Stepwise, que combina os métodos de inclusão Forward com eliminação Backward. Aplicando esta metodologia, verificou-se que em ambos os casos todas as variáveis se mostraram significativas, devendo, por esse motivo, ser incluídas em cada um dos modelos de regressão. Considerou-se como critério de escolha $\alpha = 0.05$, tanto para a inclusão como para a exclusão de variáveis. Foram ainda realizados testes de seleção de variáveis considerando como critério de escolha $\alpha = 0.01$, verificando que algumas das variáveis deixam de ser tão significativas, alertando para o facto dessas variáveis terem, possivelmente, menor importância na explicação das taxas de juro comparativamente às restantes. As variáveis seleccionadas para serem incluídas nos dois modelos estão resumidas na Tabela 3.6.

Empréstimo	Empresa
Ano	Dimensão (Total de ativo)
Montante	Setor de Atividade
Prazo Contratual	Rating
Tipo de Negociação	Crédito Vencido
Colateral	PME Líder
Bancos Top 5	Exportadora
IF Nacional vs Estrangeira	Empresa de Elevado Crescimento
	Número de Novas Operações
	Número de Instituições Financeiras

Tabela 3.6: Variáveis incluídas nos modelos

Uma vez que as variáveis explicativas respeitantes às características das empresas correspondem a dados anuais, e as variáveis respeitantes aos empréstimos correspondem a micro dados, considerando operação a operação, tornou-se necessário encontrar apenas um tipo de periodicidade de dados. Caso se considerassem todas as operações reportadas, verificar-se-ia uma repetição exaustiva das características das empresas. Suponha-se que uma empresa, no ano de 2013, tinha 200 novas operações de empréstimos, apesar das variáveis relativas à caracterização do empréstimo se alterarem, as variáveis relativas à

caracterização da empresa, nesse ano, seriam sempre as mesmas. Assim, existiria uma repetição das características da empresa, o que, no âmbito de uma análise multivariada poderia trazer problemas no que diz respeito ao impacto e à influência destas variáveis na explicação das taxas de juro. Deste modo, optou por se escolher apenas uma operação por empresa, para cada um dos anos. A decisão da escolha dessas operações foi feita considerando a nova operação de empréstimo à qual estava associada a taxa de juro mais elevada. A base de micro dados, que inicialmente continha informação relativa a 1 138 629 novas operações de empréstimo, ficou reduzida a uma amostra de cerca de 100 mil observações. É importante referir que foi realizada uma análise prévia com o intuito de encontrar a melhor alternativa de agregar e condensar a base de dados, tendo sido testadas diversas agregações de dados. No entanto, no decorrer da aplicação do modelo de regressão beta, verificou-se uma limitação computacional, pelo que só foi possível efetuar regressões para amostras até 15 mil observações. Deste modo, foi construída uma amostra estratificada, mantendo as proporções da distribuição das taxas. Por este motivo, e para fácil comparação dos dois modelos de regressão em análise, considerou-se a amostra estratificada de dimensão $n = 15\,000$ na aplicação dos dois modelos.

3.5.1 Modelo de Regressão Linear Múltipla

Como visto na Figura 3.22, o ajustamento pela distribuição Normal aos dados não é o ótimo. No entanto esse ajustamento pode ser melhorado com a aplicação de uma transformação à variável resposta, tendo-se optado pela transformação de Box-Cox.

A aplicação desta transformação à variável resposta permite reduzir a assimetria e normalizar os dados. Tal como enunciado anteriormente, a transformação de Box-Cox pressupõe a escolha do valor de λ que maximize a função de log-verosimilhança. Para determinar esse valor recorreu-se ao gráfico da Figura 3.23, onde, analisando a função de log-verosimilhança obtém-se $\lambda = 0.331313$.

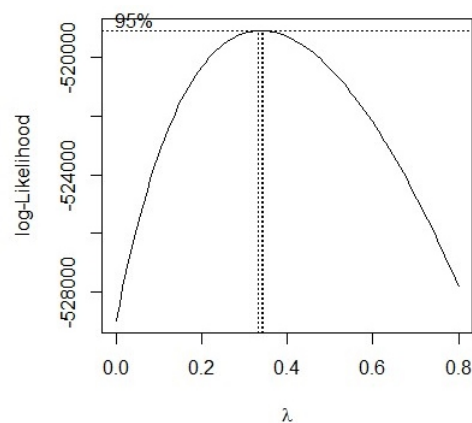


Figura 3.23: Determinação do valor de λ

Considerando este valor para λ , têm-se que $Y_i^{BC} = \frac{y_i^{0.331313} - 1}{0.331313}$.

A transformação dos dados, resultante da aplicação de Box-Cox, pode ser observada na Figura 3.24. Foi possível, através desta transformação, reduzir a assimetria dos dados.

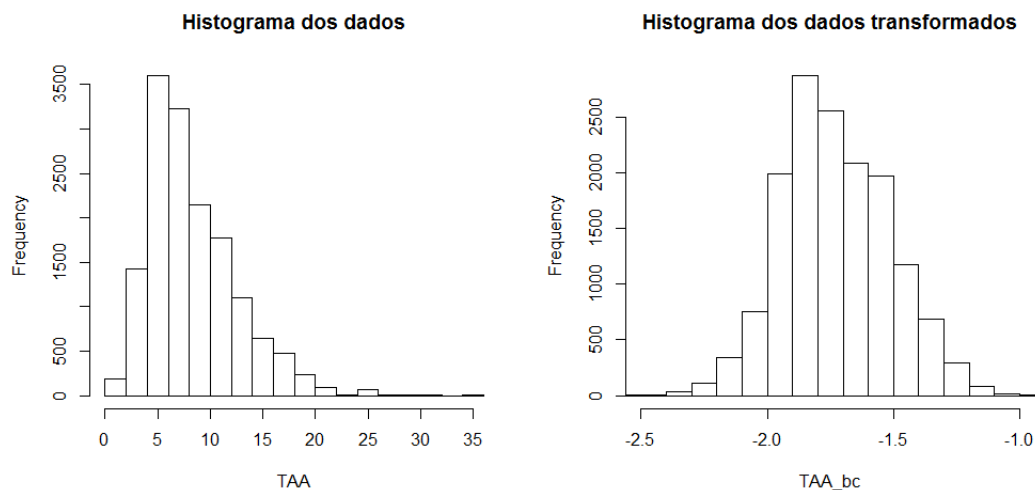


Figura 3.24: Aplicação da transformação de Box-Cox à variável resposta

Após a transformação de Box-Cox à variável resposta, foi aplicado o modelo de regressão linear múltipla.

No que diz respeito às variáveis categóricas, é importante referir a necessidade de considerar um fator padrão para cada uma delas. Quanto às variáveis binárias, por *default*, é considerado no *intercept* o valor 0. De acordo com as Tabelas 3.3 e 3.4, o fator padrão considerado para cada uma das variáveis qualitativas é resumido na Tabela 3.7.

Variável	Fator Padrão	Característica
Ano	2013	Operação realizada em 2013
Prazo Contratual	0	Prazo inferior a 3 meses
Tipo Negociação	0	Operação totalmente nova
Colateral	0	Operação sem colateral
Bancos Top 5	0	Bancos que não pertencem ao Top5
IF Nacional vs Estrangeira	0	IF Nacional
Setor de Atividade	G	Comércio
Classe de Rating	4	Classes 16-20
Crédito Vencido	0	Sem crédito vencido
PME Líder	0	Não é PME lider
Exportadora	0	Não é exportadora
EEC	0	Não é EEC

Tabela 3.7: Fatores padrão

Deste modo, os valores dos coeficientes de regressão são interpretados comparando-os com os fatores padrão.

A Tabela 3.8 fornece informação relativa às estimativas dos coeficientes de regressão, aos respetivos intervalos de confiança e ainda à significância estatística de cada regressor no modelo. A significância estatística de cada regressor é avaliada pelo teste de significância individual de Wald.

A análise da tabela permite comparar as características que levam a um agravamento das taxas de juro assim como as que originam um desagravamento das mesmas. A interpretação das estimativas dos coeficientes de regressão, pelo modelo de regressão linear múltipla, evidencia aspetos interessantes de onde se podem retirar algumas conclusões. Entre elas, destacam-se os seguintes exemplos:

- Verifica-se um desagravamento das taxas de juro dos novos empréstimos ao longo dos três anos, tal como era expectável.
- O montante, assim como o prazo contratual do empréstimo têm um impacto negativo na atribuição da taxa de juro, na medida em que quanto maior é o montante, ou o prazo contratual, mais baixa é a taxa atribuída ao empréstimo.
- Em relação ao tipo de negociação da operação registou-se que, comparativamente à taxa atribuída às operações totalmente novas (fator padrão considerado no *intercept*), as renegociações com o envolvimento do cliente têm taxas mais baixas, enquanto as renegociações sem o envolvimento do cliente ou para reestruturações de dívida têm taxas mais elevadas.
- Os bancos pertencentes ao Top 5 têm, regra geral, taxas mais baixas que a média dos bancos que não pertencem ao Top 5, excetuando o Banco D (top5_4) que verifica taxas mais elevadas comparativamente aos restantes. É de realçar que as instituições financeiras nacionais atribuem taxas mais elevadas comparativamente às taxas praticadas pelas instituições financeiras estrangeiras.
- Os coeficientes relativos ao número de novas operações e ao número de instituições financeiras sugerem a existência de um agravamento das taxas de juro quanto maior o número de novas operações da empresa assim como quanto maior o número de instituições financeiras com as quais a empresa tem novos empréstimos.
- Através dos coeficientes correspondentes ao rating da empresa, conclui-se, como se esperava, que quanto mais alta a classe de rating em que a empresa se encontra, maior a taxa de juro que lhe é atribuída.
- Verifica-se ainda que empresas PME líderes, exportadoras ou de elevado crescimento têm um desagravamento das taxas de juro.

Coeficientes	Estimativa	2.5%	97.5%	Significância
Intercept	-1.4328	-1.4669	-1.3985	***
Ano_2014	-0.0424	-0.0480	-0.0367	***
Ano_2015	-0.1216	-0.1274	-0.1158	***
Montante	-0.0296	-0.0317	-0.0276	***
PrzContr_1	-0.0795	-0.0856	-0.0733	***
PrzContr_2	-0.1331	-0.1417	-0.1246	***
PrzContr_3	-0.1378	-0.1473	-0.1284	***
OpReneg_1	-0.0002	-0.0117	0.0112	
OpReneg_2	0.1210	0.1121	0.1299	***
OpReneg_3	0.0906	0.0573	0.1239	***
Colateral_1	-0.0086	-0.0152	-0.0021	**
Top5_1	-0.0104	-0.0179	-0.0029	**
Top5_2	-0.0946	-0.1051	-0.0840	***
Top5_3	-0.0629	-0.0798	-0.0460	***
Top5_4	0.0284	0.0208	0.0359	***
Top5_5	-0.0225	-0.0316	-0.0133	***
IFnacional_1	-0.0481	-0.0643	-0.0318	***
Dimensao	-0.0248	-0.0268	-0.0226	***
Setor_A	0.0088	-0.0028	0.0204	
Setor_BC	-0.0089	-0.0154	-0.0024	**
Setor_DE	0.0067	-0.0245	0.0379	
Setor_FL	0.0025	-0.0056	0.0105	
Setor_H	0.0049	-0.0065	0.0163	
Setor_I	0.0143	0.0028	0.0258	*
Setor_JKMN	0.0011	-0.0069	0.0091	
Setor_OPQ	-0.0180	-0.0301	-0.0059	**
Setor_RSTU	0.0083	-0.0118	0.0283	
Rating_1	-0.0762	-0.1123	-0.0401	***
Rating_2	-0.0673	-0.0773	-0.0573	***
Rating_3	-0.0421	-0.0473	-0.0369	***
CredVencd_1	0.0553	0.0477	0.0628	***
PMElider_1	-0.0381	-0.0464	-0.0298	***
Exportadora_1	-0.0234	-0.0353	-0.0115	***
EEC_1	-0.0107	-0.0232	0.0018	.
NNO	0.0419	0.0388	0.0449	***
NumIF	0.0172	0.0140	0.0205	***

Tabela 3.8: Resultados do Modelo de Regressão Linear Múltipla

Na aplicação do modelo de regressão linear múltipla aos dados transformados de Box-Cox, obtiveram-se os valores de $R^2 = 0.5653$ e $\bar{R}^2 = 0.5643$. Adotando um pensamento análogo ao efetuado em Santos (2013), deve interpretar-se os resultados do modelo como

uma sistematização das características que determinam a fixação das taxas de juro. Por este motivo, a avaliação do ajustamento do modelo não deve ser apenas justificada pelo coeficiente de determinação. Esta consideração, justifica, parcialmente, que o valor não muito elevado verificado para R^2 e \bar{R}^2 não deva ter uma importância excessiva.

Aplicando o teste de significância conjunta dos regressores, obtém-se $F = 555.8$ e $p\text{-value} \simeq 0$. Este teste de significância, denominado de teste F, tem como hipótese nula $H_0 : \beta_0 = \beta_1 = \dots = \beta_p = 0$. Assim, uma vez que o $p\text{-value}$ resultante da aplicação do teste é próximo de zero, não existe evidência estatística para não rejeitar a hipótese nula, pelo que, com base na amostra considerada e para $\alpha = 0.05$, podemos inferir que pelo menos um dos β_i é diferente de zero.

Torna-se interessante, no contexto do trabalho, efetuar um teste aos submodelos, na medida em que o conjunto total de regressores pode ser dividido em dois grupos distintos. Um deles diz respeito às características dos empréstimos, enquanto que outro diz respeito às características das empresas. Considere-se:

- Modelo 1: Com todos os regressores;
- Modelo 2: Só com regressores dos empréstimos (sem regressores das empresas);
- Modelo 3: Só com regressores das empresas (sem regressores dos empréstimos).

Pretende-se, em primeiro lugar, comparar o modelo completo ao modelo apenas com regressores relativos às características dos empréstimos, ou seja, testar a hipótese nula $H_0 : \beta_j = 0, \forall j \in \{\text{regressores relativos às características das empresas}\}$, pretendendo, deste modo, verificar se as variáveis relativas às características das empresas são significativas para o modelo. O valor observado da estatística de teste foi $F = 239.65$, com o respetivo $p\text{-value}$ muito próximo de zero, não existindo evidência para não rejeitar a hipótese nula. Conclui-se assim que, para $\alpha = 0.05$ as variáveis respeitantes às características das empresas são importantes na explicação da variável resposta, devendo, por esse motivo, ser incluídas no modelo

De forma similar, realizou-se a comparação do modelo completo com o modelo apenas com os regressores relativos às características das empresas. A hipótese nula é $H_0 : \beta_j = 0, \forall j \in \{\text{regressores relativos às características dos empréstimos}\}$, onde se pretende averiguar se as características relativas aos empréstimos são conjuntamente significativas. Observou-se que o valor da estatística de teste foi $F = 511.02$, cujo $p\text{-value}$ associado é, novamente, muito próximo de zero. Conclui-se que as variáveis relativas aos empréstimos são igualmente importantes, pelo que também devem permanecer no modelo.

Os testes aos submodelos rejeitaram as hipóteses de nulidade dos grupos de variáveis consideradas. Tanto as variáveis relativas aos empréstimos como as relativas às empresas

são, em conjunto, importantes na explicação da variável resposta, pelo que devem pertencer ao modelo de regressão. Estes resultados eram os esperados tendo em conta os níveis de significância indicados para cada variável. como indicado na Tabela 3.8. Um vez que a dimensão da amostra é bastante elevada, deve ter-se alguma atenção à explicação do baixo valor do p -value. No Capítulo 5, referente à influência da dimensão das amostras, será apresentado um estudo mais detalhado acerca da influência que as grandes amostras podem ter nos resultados estatísticos relativos ao ajustamento das regressões.

A análise dos resíduos do modelo, permite avaliar a adequabilidade do modelo. Deste modo, é possível detetar a existência, ou não, de tendência entre os resíduos, o que indica um melhor, ou pior, ajustamento do modelo de regressão. Através da análise da Figura 3.25 observa-se que os resíduos concentram-se em torno de zero e não apresentam qualquer tendência. Ambos os gráficos sugerem a ausência de dependência dos resíduos, sendo este um fator indicativo da adequabilidade do modelo.

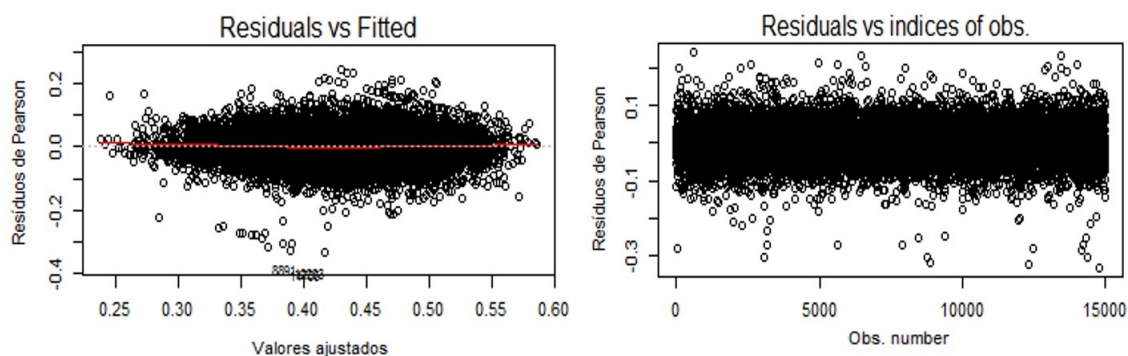


Figura 3.25: Resíduos de Pearson em função dos valores estimados e das observações no modelo de regressão linear múltipla

A distância de Cook, é uma medida de análise de resíduos que permite identificar possíveis *outliers*. Quanto maior o valor da distância de Cook, maior o impacto que a exclusão de determinada observação implica no ajustamento da regressão. Através da análise da Figura 3.26, observa-se que, a grande maioria das observações têm distâncias de Cook inferiores a 0.0005, verificando-se, no entanto, algumas observações acima de 0.002, sendo estas distâncias consideradas relativamente baixas.

De um modo geral, o modelo de regressão linear múltipla, com a transformação de Box-Cox à variável resposta, parece ajustar-se bem aos dados em análise. Existe evidência de que este modelo seja adequado na compreensão dos fatores determinantes do nível das taxas de juro bancárias relativas a novas operações de empréstimos a sociedades não financeiras.

3.5.2 Modelo de Regressão Beta

Um modelo alternativo ao apresentado anteriormente é o modelo de regressão beta. Como explicado no Capítulo 2, a aplicação do modelo de regressão beta poderá ser adequada

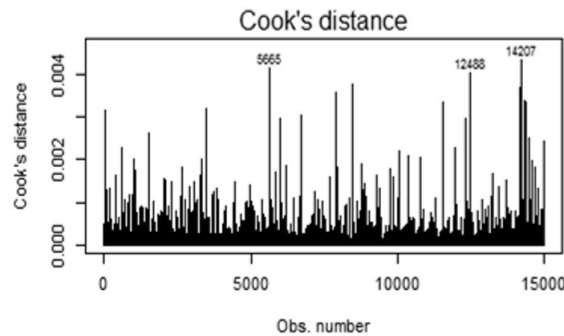


Figura 3.26: Distâncias de Cook no modelo de regressão linear múltipla

aos casos em que a variável resposta é contínua e pertence ao intervalo $(0, 1)$. Deste modo, e uma vez que a variável que se pretende explicar diz respeito às taxas de juro bancárias, faz sentido a aplicação deste modelo, sendo que se trata de uma variável contínua, com valores entre 0 e 1.

O modelo de regressão beta pressupõe a escolha de uma função de ligação. Ajustou-se mais que um modelo de regressão beta, fazendo alterar a função ligação. O melhor ajustamento foi conseguido utilizando a função de ligação log-log, Tabela 3.9, dada por $g(\mu) = -\log(-\log(\mu))$. As medidas utilizadas para avaliar qual a função ligação mais apropriada foram o valor do pseudo coeficiente de determinação e o valor do AIC (Critério de Informação de Akaike) que mede a qualidade relativa de um modelo estatístico.

Função de Ligação	Pseudo R^2	AIC
Logit	0,5377	-67360,62
Probit	0,5551	-67346,63
Log-Log	0,5585	-67396,93
Log-Log complementar	0,5349	-66658,41
Cauchy	0,0115	-67347,71

Tabela 3.9: Escolha da função de ligação

Decidida a função de ligação a utilizar, procedeu-se ao ajustamento do modelo de regressão beta. Os fatores considerados como fatores padrão, são os mesmos que foram anteriormente considerados para a aplicação do modelo de regressão linear, Tabela 3.7.

Os resultados provenientes da aplicação do modelo de regressão beta aos dados, estão presentes na Tabela 3.10, onde são indicadas as estimativas dos coeficientes de regressão, os intervalos de confiança associados, assim como as respetivas significâncias estatísticas resultantes do teste de significância individual de Wald. A interpretação dos coeficientes estimados pelo modelo de regressão beta, permite retirar conclusões muito semelhantes às da aplicação do modelo de regressão linear múltipla.

Coeficientes	Estimativa	2.5%	97.5%	Significância
Intercept	-0.6554	-0.6865	-0.6243	***
Ano_2014	-0.0375	-0.0426	-0.0324	***
Ano_2015	-0.1083	-0.1135	-0.1029	***
Montante	-0.0256	-0.0275	-0.0237	***
PrzContr_1	-0.0730	-0.0785	-0.0675	***
PrzContr_2	-0.1186	-0.1264	-0.1108	***
PrzContr_3	-0.1234	-0.1322	-0.1147	***
OpReneg_1	0.0005	-0.0102	0.0111	
OpReneg_2	0.1081	0.1000	0.1160	***
OpReneg_3	0.0771	0.0468	0.1074	***
Colateral_1	-0.0081	-0.0139	-0.0021	**
Top5_1	-0.0094	-0.0162	-0.0026	**
Top5_2	-0.0837	-0.0935	-0.0738	***
Top5_3	-0.0434	-0.0594	-0.0274	***
Top5_4	0.0283	0.0216	0.0351	***
Top5_5	-0.0165	-0.0248	-0.0083	***
IFnacional_1	-0.0530	-0.0684	-0.0377	***
Dimensao	-0.0223	-0.0243	-0.0204	***
Setor_A	0.0079	-0.0027	0.0185	
Setor_BC	-0.0074	-0.0133	-0.0015	*
Setor_DE	0.0037	-0.0253	0.0326	
Setor_FL	0.0032	-0.0041	0.0105	
Setor_H	0.0044	-0.0060	0.0148	
Setor_I	0.0122	0.0018	0.0226	*
Setor_JKMN	0.0008	-0.0065	0.0080	
Setor_OPQ	-0.0170	-0.0282	-0.0058	**
Setor_RSTU	0.0080	-0.0102	0.0262	
Rating_1	-0.0673	-0.1013	-0.0334	***
Rating_2	-0.0599	-0.0693	-0.0506	***
Rating_3	-0.0375	-0.0422	-0.0328	***
CredVencd_1	0.0500	0.0433	0.0567	***
PMElider_1	-0.0328	-0.0404	-0.0251	***
Exportadora_1	-0.0219	-0.0329	-0.0109	***
EEC_1	-0.0086	-0.0201	0.0028	.
NNO	0.0372	0.0345	0.0400	***
NumIF	0.0158	0.0129	0.0187	***

Tabela 3.10: Resultados do Modelo de Regressão Beta

Através da análise da Tabela 3.10 é possível observar os seguintes aspetos:

- O aumento do montante dos empréstimos tem um efeito negativo, verificando-se uma diminuição das taxas de juro praticadas.

- Existe uma relação inversa entre as taxas de juro e o prazo contratual, na medida em que quanto maior for a maturidade dos empréstimos menor é a taxa de juro atribuída.
- As empresas PME líderes, exportadoras e de elevado crescimento têm um desagravamento das taxas de juro, tal como também foi possível verificar através da regressão linear múltipla.
- Como se esperava, a dimensão da empresa e a taxa que lhe é atribuída têm uma relação inversa, uma vez que empresas de grandes dimensões têm taxas mais baixas comparativamente a pequenas empresas.
- A classe de rating, confirma a importância da situação financeira das empresas na determinação das respetivas taxas de juro, isto porque são atribuídas taxas mais elevadas a empresas com maior risco.

Duas das medidas comuns do ajustamento to modelo de regressão beta são o *pseudo* R^2 e o *AIC*, informações presentes na Tabela 3.9. Verifica-se que *pseudo* $R^2 = 0.5585$ e que $AIC = -67356.93$, no entanto este último por si só não é ilustrativo de um bom ajustamento, apenas quando comparado com outros valores, como na tabela 3.9.

Com o intuito de avaliar a significância global do modelo, foi efetuado um teste de significância conjunta de todos os regressores. Para tal, recorreu-se ao teste de razão de verosimilhanças, tal como sugerido por Cribari-Neto e Zeileis (2009). A estatística de teste obtida foi $w_1 = 12035$, que origina um *p*-value aproximadamente igual a zero, sugerindo que os regressores sejam conjuntamente significativos.

Com o intuito de analisar a significância do grupo de características da empresa e do grupo de características do empréstimo, foram realizados, também para o modelo de regressão beta, testes aos submodelos. Esta análise, à semelhança da efetuada no modelo de regressão simples, permite avaliar a importância, ou não, de cada um destes conjuntos de regressores. Os testes aos submodelos foram realizados recorrendo ao teste de razão de verosimilhanças.

Ao comparar o modelo completo com o modelo apenas com regressores relativos às características do empréstimo, pretende-se testar a significância do conjunto de regressores respeitantes às características das empresas. O valor observado da estatística de teste foi $w_1 = 2669.1$, o que evidencia, a um nível de significância de $\alpha = 0.05$, a importância dos regressores relativos às características da empresa para o modelo. Do mesmo modo, ao efetuar a comparação do modelo completo com o modelo apenas com os regressores relativos às características das empresas, pretende-se avaliar a significância conjunta dos regressores relativos às características dos empréstimos. O valor respeitante à estatística de teste foi $w_1 = 4388.9$, sendo indicativo de que o ajustamento do modelo melhora quando se consideram também as características dos empréstimos.

À semelhança do que foi referido anteriormente para o modelo de regressão linear, também no que diz respeito ao modelo de regressão beta será feito um estudo relativo ao comportamento do p -value para amostras de grande dimensão no Capítulo 5.

Uma análise baseada nos resíduos gerados pelo modelo de regressão beta, permite concluir que estes não apresentam tendência nem dependência entre si, e que se distribuem em torno de zero, tal como se pode observar na Figura 3.27.

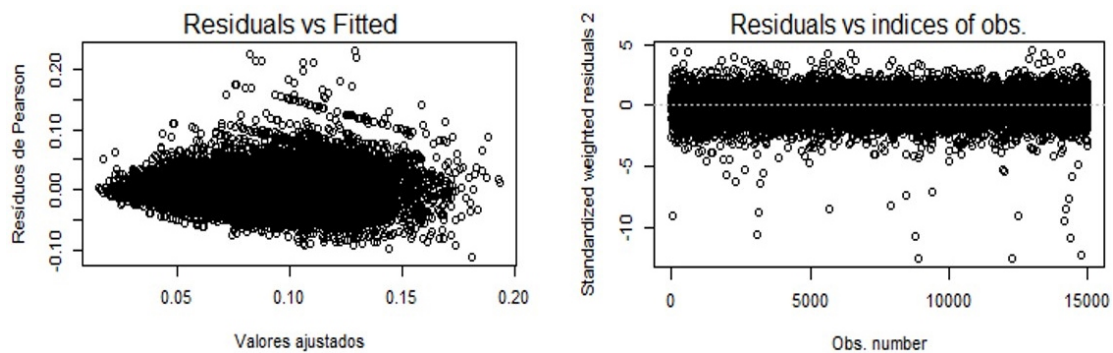


Figura 3.27: Resíduos de Pearson e resíduos padronizados ponderados aplicados ao modelo de regressão beta

O primeiro gráfico corresponde aos resíduos de Pearson em função dos valores ajustados, enquanto que o segundo gráfico diz respeito aos resíduos padronizados ponderados, sugeridos por Cribari-Neto e Zeileis (2009), em função do número de observações. A análise das distâncias de Cook, permite evidenciar a elevada concentração de distâncias em valores abaixo dos 0.0005, sendo indicativo de poucas observações consideradas influentes. A análise gráfica, relativa às distâncias de Cook aplicadas ao modelo de regressão beta, está presente na Figura 3.28.

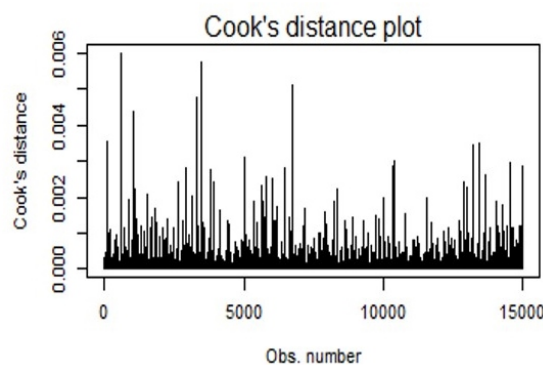


Figura 3.28: Distâncias de Cook aplicadas ao modelo de regressão beta

3.5.3 Comparação dos Modelos

Os dois modelos apresentados anteriormente, modelo de regressão linear e modelo de regressão beta, pretendem explicar o comportamento das taxas de juro de novos empréstimos, através de um conjunto de características dos empréstimos e das empresas.

Coeficientes	Modelo de Regressão Linear		Modelo de Regressão Beta	
	Estimativa	Significância	Estimativa	Significância
Intercept	-1.4328	***	-0.6554	***
Ano_2014	-0.0424	***	-0.0375	***
Ano_2015	-0.1216	***	-0.1083	***
Montante	-0.0296	***	-0.0256	***
PrzContr_1	-0.0795	***	-0.0730	***
PrzContr_2	-0.1331	***	-0.1186	***
PrzContr_3	-0.1378	***	-0.1234	***
OpReneg_1	-0.0002		0.0005	
OpReneg_2	0.1210	***	0.1081	***
OpReneg_3	0.0906	***	0.0771	***
Colateral_1	-0.0086	**	-0.0081	**
Top5_1	-0.0104	**	-0.0094	**
Top5_2	-0.0946	***	-0.0837	***
Top5_3	-0.0629	***	-0.0434	***
Top5_4	0.0284	***	0.0216	***
Top5_5	-0.0225	***	-0.0165	***
IFnacional_1	-0.0481	***	-0.0530	***
Dimensao	-0.0248	***	-0.0223	***
Setor_A	0.0088		0.0079	
Setor_BC	-0.0089	**	-0.0074	*
Setor_DE	0.0067		0.0037	
Setor_FL	0.0025		0.0032	
Setor_H	0.0049		0.0044	
Setor_I	0.0143	*	0.0122	*
Setor_JKMN	0.0011		0.0008	
Setor_OPQ	-0.0180	**	-0.0170	**
Setor_RSTU	0.0083		0.0080	
Rating_1	-0.0762	***	-0.0673	***
Rating_2	-0.0673	***	-0.0599	***
Rating_3	-0.0421	***	-0.0375	***
CredVencd_1	0.0553	***	0.0500	***
PMElider_1	-0.0381	***	-0.0328	***
Exportadora_1	-0.0234	***	-0.0219	***
EEC_1	-0.0107	.	-0.0086	.
NNO	0.0419	***	0.0372	***
NumIF	0.0172	***	0.0158	***

Tabela 3.11: Comparação das estimativas dos coeficientes pelos dois modelos de regressão

Verificou-se, através da análise multivariada efetuada nas Subsecções 3.5.1 e 3.5.2, um bom ajustamento dos modelos, tendo sido possível compreender e interpretar o contributo positivo ou negativo de cada regressor na explicação da determinação das taxas de juro em novos empréstimos.

Os sinais das estimativas dos coeficientes de regressão no modelo linear e no modelo beta, tal como se pode concluir através da análise da Tabela 3.11, são, na sua maioria, coincidentes, pelo que podemos afirmar que existe concordância nos fatores considerados agravantes e desagravantes da fixação das taxas de juro em ambos os modelos.

A Figura 3.29 compara a densidade empírica das taxas de juro observadas com as densidades das taxas ajustadas pelos dois modelos de regressão em estudo.

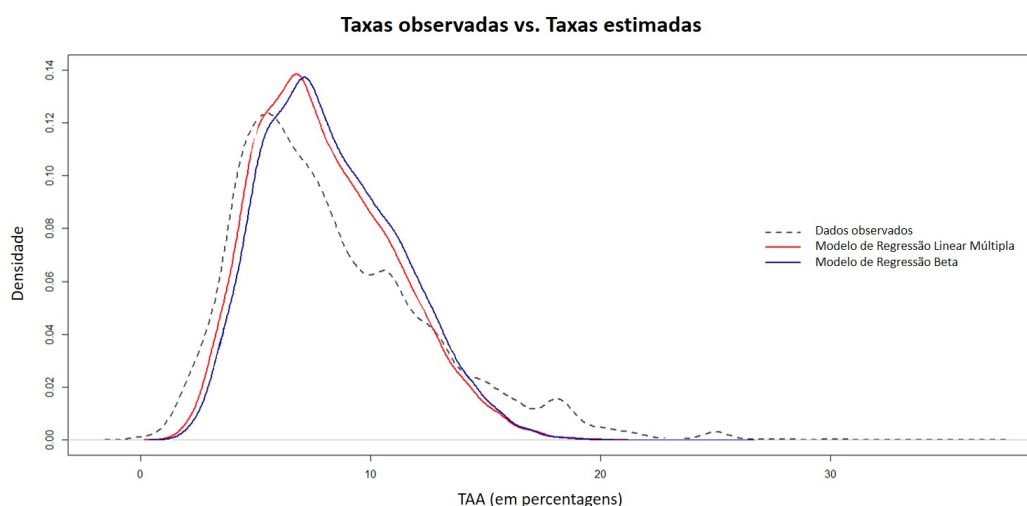


Figura 3.29: Comparação das densidades empírica e teóricas relativas às taxas de juro

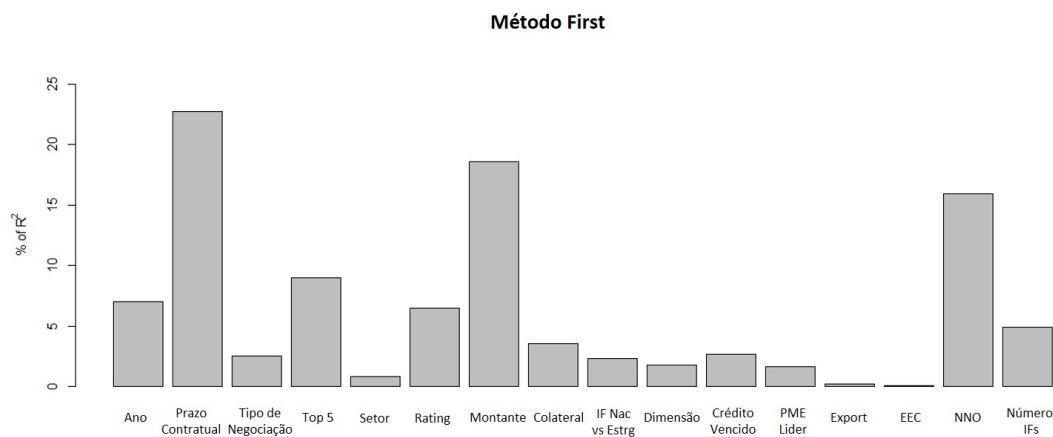
Através desta figura, é possível observar uma maior concentração das taxas até aos 10%, tanto nas taxas de juro observadas como nas estimadas pelas duas regressões.

ANÁLISE DE RESULTADOS

4.1 Importância Relativa dos Regressores

As taxas de juro são influenciadas por um conjunto de fatores, como foi possível concluir pela análise multivariada efetuada no Capítulo 3. No entanto, mais do que entender o agravamento ou desagravamento que cada fator tem na fixação das taxas de juro, torna-se interessante determinar qual a importância que cada uma das variáveis tem para o modelo. De acordo com a teoria apresentada na Secção 2.1.2, e recorrendo ao package 'Relaimpo', disponível para o software R, é possível dar resposta a esta questão.

Para o presente trabalho, analisaram-se duas das quatro métricas simples e duas métricas mais complexas, que exigem um maior esforço computacional, sendo elas as métricas First, Last, LMG e PMVD. É importante referir que esta metodologia de avaliação da importância dos regressores apenas pode ser aplicada em modelos de regressão linear. Na Figura 4.1 são apresentados os gráficos relativos à aplicação das quatro métricas referidas.



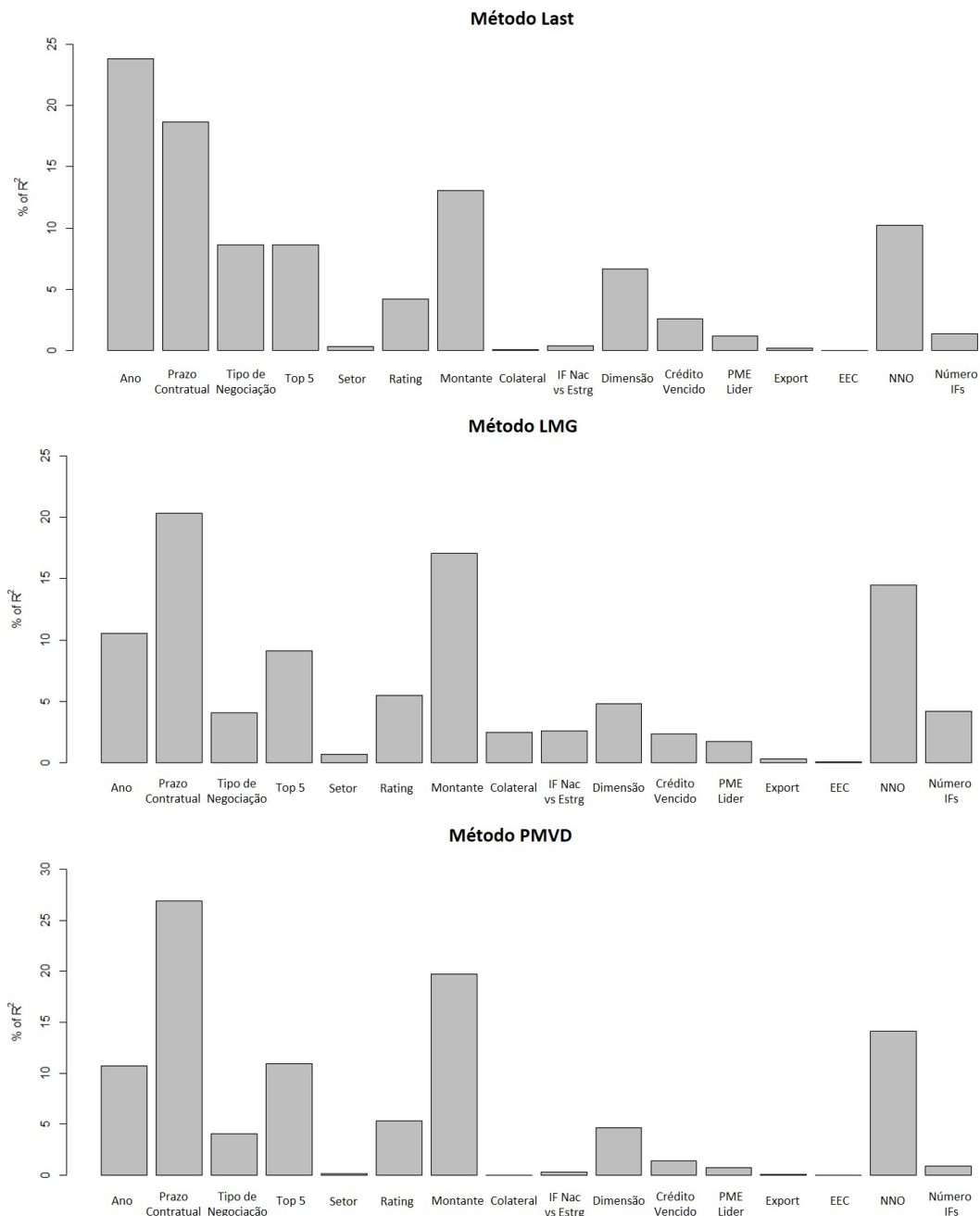


Figura 4.1: Comparação dos métodos First, Last, LMG e PMVD na avaliação da importância das variáveis em estudo

Comparando as duas métricas simples, First e Last, verificam-se algumas diferenças significativas nas percentagens relativas às contribuições dos regressores, nomeadamente nas variáveis relativas ao ano do empréstimo, ao tipo de negociação e à dimensão da empresa. No que toca à comparação dos métodos LMG e PMVD, não se observam diferenças significativas, no entanto, as variáveis com contribuições mais reduzidas têm percentagens mais baixas quando se utiliza o método PMVD comparativamente ao método LMG.

4.1. IMPORTÂNCIA RELATIVA DOS REGRESSORES

Através dos diferentes métodos de medição da contribuição individual dos regressores, é possível encontrar um conjunto de regressores com um peso elevado na atribuição das taxas de juro. Entre esses regressores, salientam-se como características dos empréstimos o ano, o prazo contratual, o montante e ainda o facto do empréstimo ter sido realizado num dos bancos pertencentes ao Top 5. No que diz respeito às características das empresas, as que têm maior peso são a dimensão e a classe de rating.

De seguida será apresentado o exemplo de uma operação escolhida aleatoriamente da base de dados, para a qual se efetuará uma análise detalhada dos impactos que determinadas alterações nas características do empréstimo e da empresa teriam para o valor da taxa de juro atribuída.

A empresa que efetuou o empréstimo analisado será denominada, por questões de confidencialidade, de "Empresa ABC", apresentando as características da Tabela 4.1.

Empréstimo da empresa ABC	
Ano do Empréstimo	2015
Montante	11 203 €
Prazo Contratual	2 - de 1 a 5 anos
Tipo de Negociação	0 - Operação totalmente nova
Colateral	1 - Com colateral
Bancos Top5	0 - Outros bancos
IF Nacional vs Estrangeira	1 - Estrangeira
Total de Ativo (Dimensão)	1.6 Milhões € (Pequena Empresa)
Setor de Atividade	BC - Indústrias
Rating	2 - Classes 6-10
Crédito Vencido	0 - Sem crédito vencido
PME Líder	1 - É PME líder
Exportadora	0 - Não é exportadora
EEC	0 - Não é EEC
NNO	19
Número de IF's	3

Tabela 4.1: Características do empréstimo relativo à empresa ABC

A taxa de juro atribuída a esta operação foi de 5.0116%, sendo que, para esta mesma operação, as taxas de juro estimadas foram de 4.9294% pelo modelo de regressão linear múltipla e de 5.0921% pelo modelo de regressão beta.

Com o objetivo de avaliar o impacto que alterações nas características dos empréstimos têm na atribuição das taxas de juro, procedeu-se a uma análise de sensibilidade, sendo possível confirmar a importância dos regressores identificados anteriormente. A Tabela 4.2 indica alguns exemplos onde se observam variações das taxas de juro resultantes de alterações das características dos empréstimos.

Característica	MR Linear	MR Beta
	4.9294%	5.0921%
Ano		
2015 → 2013	6.7395% +36.7%	6.9114% +35.7%
Montante		
11 203€ → 1 000 000€	3.3562% -31.9%	3.5426% -30.4%
11 203€ → 1 000€	5.9499% +20.7%	6.0877% +19.5%
Prazo Contratual		
2 → 0	6.9313% +40.6%	7.1043% +36.5%
Sector de Atividade		
BC → A	5.1695% +4.9%	5.3269% +4.6%
BC → OPQ	4.80849% -2.5%	4.9470% -2.8%
Classe de Rating		
2 → 4	5.8844% +19.4%	6.0552% +18.9%
Exportadora		
0 → 1	4.6233% -6.2%	4.7664% -6.4%
NNO		
19 → 100	5.9183% +20.1%	6.0881% +19.6%

Tabela 4.2: Variações nas taxas de juro

Para uma fácil compreensão da tabela, veja-se o exemplo de uma variação no ano do empréstimo. Segundo as características do empréstimo da empresa ABC, presentes na Tabela 4.1, obtêm-se, pelos dois modelos de regressão em estudo as taxas referidas na primeira linha da tabela, 4.9294% e 5.0921%. O empréstimo em análise data de 2015, mas, caso tivesse ocorrido em 2013, a taxa estimada pelo modelo linear seria de 6.7395% e pelo modelo beta seria de 6.9114%. Assim, para uma operação com exatamente as mesmas características excetuando o ano em que fora realizada, obter-se-ia um aumento de 36.7% e de 35.7% nas taxas estimadas pelo modelo linear e pelo modelo beta, respetivamente.

Pela análise da tabela anterior, é possível confirmar que o ano de realização do empréstimo, o montante e o prazo contratual são variáveis com uma elevada contribuição para a explicação da variável resposta, taxas de juro das novas operações de empréstimos. A classe de rating, tal como visto anteriormente é também uma variável com grande

importância enquanto fator determinante da fixação das taxas de juro de novas operações de empréstimos. O facto da empresa ser ou não exportadora tem impacto na variação das taxas de juro, no entanto essa variação é mais reduzida comparativamente com a variação das restantes variáveis consideradas. É possível observar que um aumento do montante do empréstimo faz diminuir as taxas de juro, enquanto que uma diminuição do montante leva a um aumento das taxas.

O prazo contratual considerado no empréstimo em análise pertence ao intervalo temporal de 1 a 5 anos (prazo contratual=2), mas, considerando um prazo até 3 meses (prazo contratual=0), as taxas de juro aumentam bastante. O mesmo acontece com a classe de rating da empresa. A empresa em análise tem uma avaliação de rating entre as classe 6 e 10 (rating=2), no entanto, se considerarmos uma classificação de rating elevada, entre as classes 16 e 20 (rating=4), as taxas de juro do empréstimo aumentam significativamente.

No que toca ao setor de atividade, verifica-se que, comparativamente ao setor das indústrias (setor BC), existe um aumento das taxas de juro quando, mantendo tudo o resto constante, se considera um empréstimo feito por uma empresa do setor da agricultura (setor A). Verifica-se, por exemplo, uma diminuição das taxas de juro no caso do empréstimo ser realizado por uma empresa pertencente ao setor da Administração Pública, Saúde, Educação e Segurança Social (setor OPQ).

Segundo Grömping et al. (2006) e como explicado na secção 2.1.2, o método que proporciona uma melhor e mais correta explicação das contribuições individuais de cada um dos regressores é o método PMVD. Neste método são consideradas todas as ordenações possíveis de inclusão das variáveis no modelo, proporcionando diferentes contribuições no R^2 . Por este motivo, nas análises seguintes, serão apresentados apenas os resultados obtidos através da aplicação deste método.

4.2 Comparação das Instituições Financeiras Nacionais e Estrangeiras

Uma possível análise relativa às taxas de juro baseia-se na comparação das taxas praticadas pelas instituições financeiras nacionais comparativamente às taxas praticadas pelas instituições financeiras estrangeiras. Tal como visto nas estimativas dos coeficientes das duas regressões estudadas, existe evidência de que as instituições financeiras nacionais praticam, em média, taxas mais elevadas que as instituições financeiras estrangeiras.

A Figura 4.2 ilustra as densidades das taxas estimadas pelo modelo de regressão linear simples analisando as taxas de juro relativas aos empréstimos realizados em IF's nacionais comparativamente às taxas relativas aos empréstimos realizados em IF's

estrangeiras. Verifica-se uma elavada concentração de taxas em torno dos 5% nos empréstimos realizados em IF's estrangeiras. As taxas relativas a empréstimos realizados em instituições financeiras nacionais concentram-se num intervalo mais alargado, entre os 5% e os 13%.

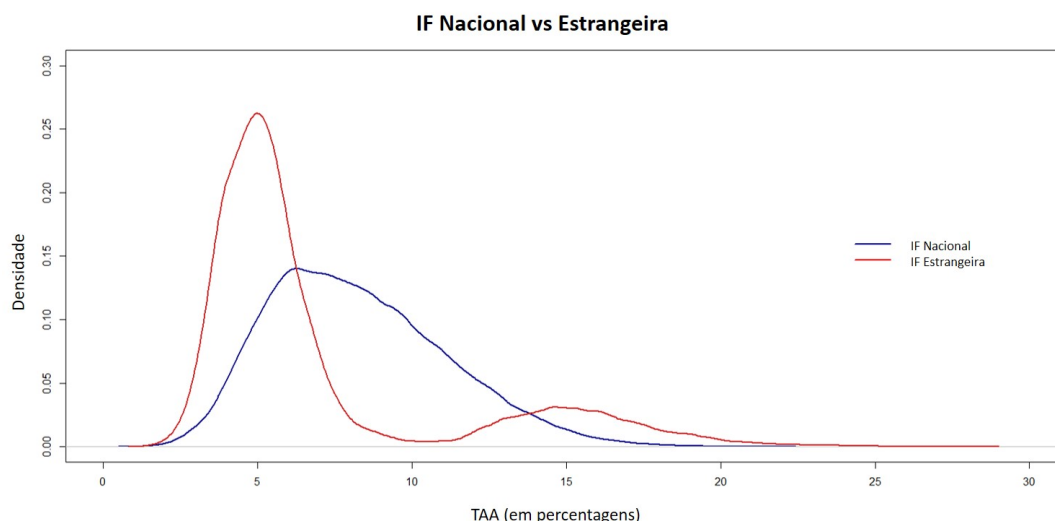


Figura 4.2: Densidade das taxas de juro comparando IF's nacionais e estrangeiras

Na Tabela 4.3 são indicadas as variações nas taxas de juro quando se consideram os empréstimos relativos a todas as instituições financeiras nacionais comparativamente às taxas praticadas pelas instituições financeiras estrangeiras. O empréstimo tido como exemplo no presente trabalho, cujas características estão presentes na Tabela 4.1, foi realizado numa instituição financeira estrangeira (IFnacional=1) que não pertence ao Top 5 (Top5=0). Na comparação dos dois tipos de IF's são consideradas todas as instituições financeiras nacionais e todas as instituições financeiras estrangeiras.

Característica	MR Linear	MR Beta
	4.9294%	5.0921%
Instituição Financeira		
Nacional	5.5196%	5.9085%
	+11.9%	+16.0%
Estrangeira	4.3157%	4.6260%
	-12.5%	-9.2%

Tabela 4.3: Variações nas taxas de juro comparando IF's nacionais e estrangeiras

Pela análise da tabela, verifica-se, em média, um aumento das taxas de juro quando se consideram todas as IF's nacionais. Observa-se ainda que quando se consideram todas as IF's estrangeiras, pertencentes e não pertencentes ao Top 5, as taxas estimadas diminuem.

4.2. COMPARAÇÃO DAS INSTITUIÇÕES FINANCEIRAS NACIONAIS E ESTRANGEIRAS

A comparação dos fatores determinantes na fixação das taxas de juro por IF's nacionais e estrangeiras pode ser realizada através da medição das contribuições de cada regressor no modelo. A Figura 4.3 permite identificar diferenças significativas nas características mais importantes na fixação das taxas de juro, consideradas pelos dois tipos de IF's.

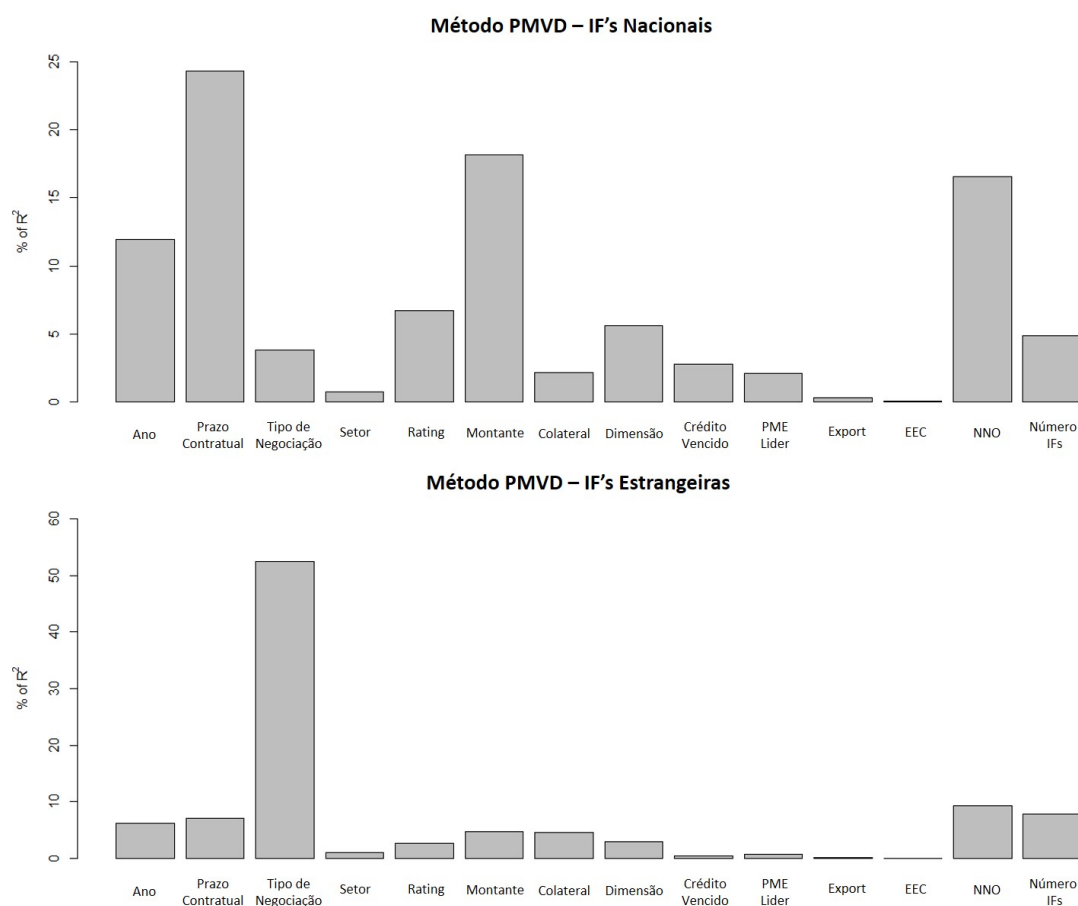


Figura 4.3: Importância dos regressores comparando IF's nacionais e estrangeiras

Ao analisar as variáveis mais influentes na atribuição das taxas de juro consoante a nacionalidade da instituição financeira, parece verificar-se uma diferença significativa na importância que as IF's dão ao tipo de negociação do empréstimo. Ou seja, parece existir uma elevada importância dada, por parte das IF's estrangeiras, ao facto da operação ser totalmente nova ou ser renegociada.

Relativamente às restantes variáveis parece existir uma concordância nas suas contribuições para a explicação do modelo em estudo. Os regressores com maior contribuição para o modelo são o ano, o montante, a maturidade do empréstimo, assim como o número de novas operações e o número de instituições financeiras com as quais a empresa tem novas operações de empréstimos.

ANÁLISE DA VARIAÇÃO DOS RESULTADOS EM FUNÇÃO DA DIMENSÃO DAS AMOSTRAS

No presente trabalho, como referido anteriormente, a base de dados utilizada tem uma dimensão bastante elevada, fator que pode condicionar os resultados obtidos. Por esse motivo, foi aplicada uma metodologia de avaliação do impacto que amostras de grande dimensão têm nos resultados obtidos pelos modelos de regressão. Esta análise tem como objetivo confirmar que os resultados estatísticos obtidos não são enganosos e ineficazes, nem influenciados pela dimensão da amostra.

De acordo com o exposto na Secção 2.4 e sugerido por Lin et al. (2013), serão aplicadas metodologias de análise da influência da dimensão das amostras nos resultados estatísticos. As metodologias adotadas passam pela construção de amostras de menor dimensão, para as quais são aplicados os modelos de regressão. Deste modo, são obtidas as estimativas dos coeficientes da regressão, os respetivos intervalos de confiança e os p -values associados, como função das dimensões das amostras construídas.

5.1 Modelo de Regressão Linear Múltipla

Na presente secção será abordado o eventual problema das grandes amostras para os resultados do modelo de regressão linear múltipla com a aplicação da transformação de Box-Cox. Será apresentada a análise do impacto da dimensão da amostra numa das variáveis explicativas do modelo, a variável relativa ao montante das novas operações de empréstimos. A escolha desta variável deveu-se ao facto de, tal como verificado no Capítulo 4, ser uma das variáveis mais importantes na fixação das taxas de juro de novas operações de empréstimos bancários. No entanto, foi realizada uma análise muito semelhante para as restantes variáveis. Será ainda avaliado o impacto da dimensão da

CAPÍTULO 5. ANÁLISE DA VARIAÇÃO DOS RESULTADOS EM FUNÇÃO DA DIMENSÃO DAS AMOSTRAS

amostra nas medidas relativas ao ajustamento do modelo, nomeadamente o valor do R^2 , do \bar{R}^2 e do p -value resultante do teste F que avalia a significância global do modelo.

A Figura 5.1 apresenta a estimação do coeficiente de regressão e o respetivo intervalo de confiança da variável relativa ao montante do empréstimo, fazendo aumentar a dimensão da amostra. Esta figura apresenta ainda o valor do p -value, com o intuito de mostrar de que forma a dimensão da amostra afeta este valor.

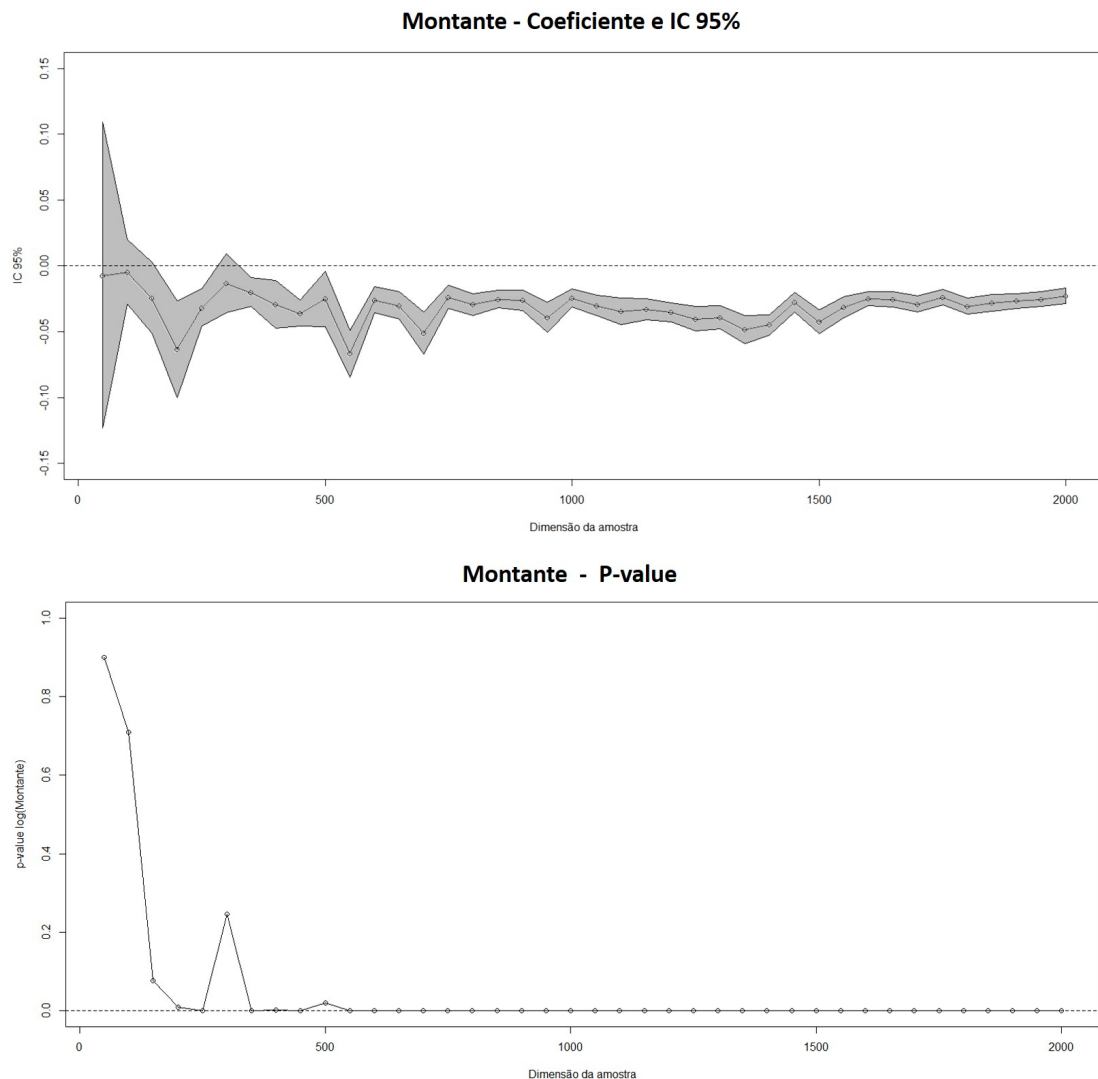


Figura 5.1: Coeficiente, intervalo de confiança a 95% e p -value associados à variável Montante através do modelo de regressão linear

Na Figura 5.1 pode verificar-se que as estimativas dos coeficientes de regressão estabilizam para dimensões da amostra relativamente baixas. O mesmo acontece para o p -value que rapidamente se aproxima de zero. Ao realizar uma análise semelhante para as restantes variáveis, verificou-se que a partir de uma amostra de dimensão $n = 1200$ a grande maioria das variáveis se tornam significativas a 1%. Comparando

o comportamento descrito nos gráficos anteriores, pode observar-se que a dimensão da amostra parece afetar mais o comportamento do p -value comparativamente ao da estimação dos coeficientes e dos respetivos intervalos de confiança.

Os gráficos da Figura 5.1 baseiam-se apenas numa única reamostragem para cada dimensão da amostra, sendo, no entanto, interessante realizar várias reamostragens para as diferentes dimensões da amostra. Recorreu-se a 1000 reamostragens para cada uma das dimensões da amostra, com o objetivo de estudar o p -value como função desta última.

Com base nestas reamostragens podem ser analisados os coeficientes estimados, e os respetivos p -values associados, assim como medidas estatísticas relativas ao ajustamento do modelo. No presente trabalho optou-se por analisar o comportamento do p -value resultante do teste de significância conjunta do modelo, assim como o comportamento do R^2 e do \bar{R}^2 .

A Figura 5.2 corresponde à representação gráfica dos valores do p -value relativos ao teste F , que avalia a significância conjunta dos regressores. Verifica-se não só que o valor do p -value decresce muito rapidamente à medida que a dimensão da amostra aumenta, mas também que a respetiva variabilidade reduz. No entanto, para amostras muito pequenas, o p -value toma, desde logo, valores muito próximos de zero.

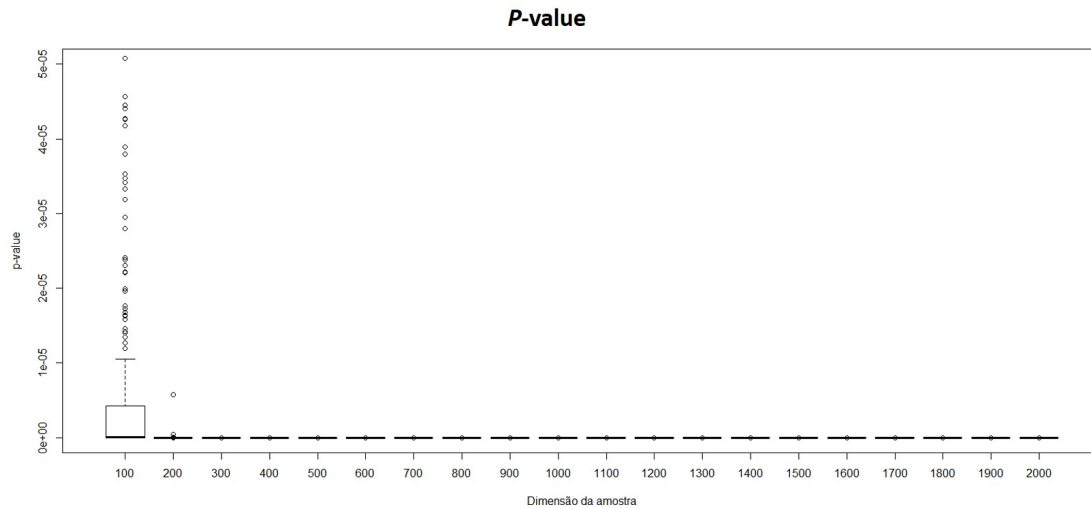


Figura 5.2: Análise do p -value do teste F

Como referido anteriormente, é possível analisar o comportamento do R^2 e do \bar{R}^2 em função da dimensão da amostra. A Figura 5.3 reflete a variação destes dois coeficientes de determinação.

Pela análise dos resultados, observa-se que à medida que a dimensão da amostra aumenta, os valores do R^2 e do \bar{R}^2 tendem a estabilizar, verificando-se uma diminuição das suas variabilidades. É importante realçar que, para amostras muito pequenas os

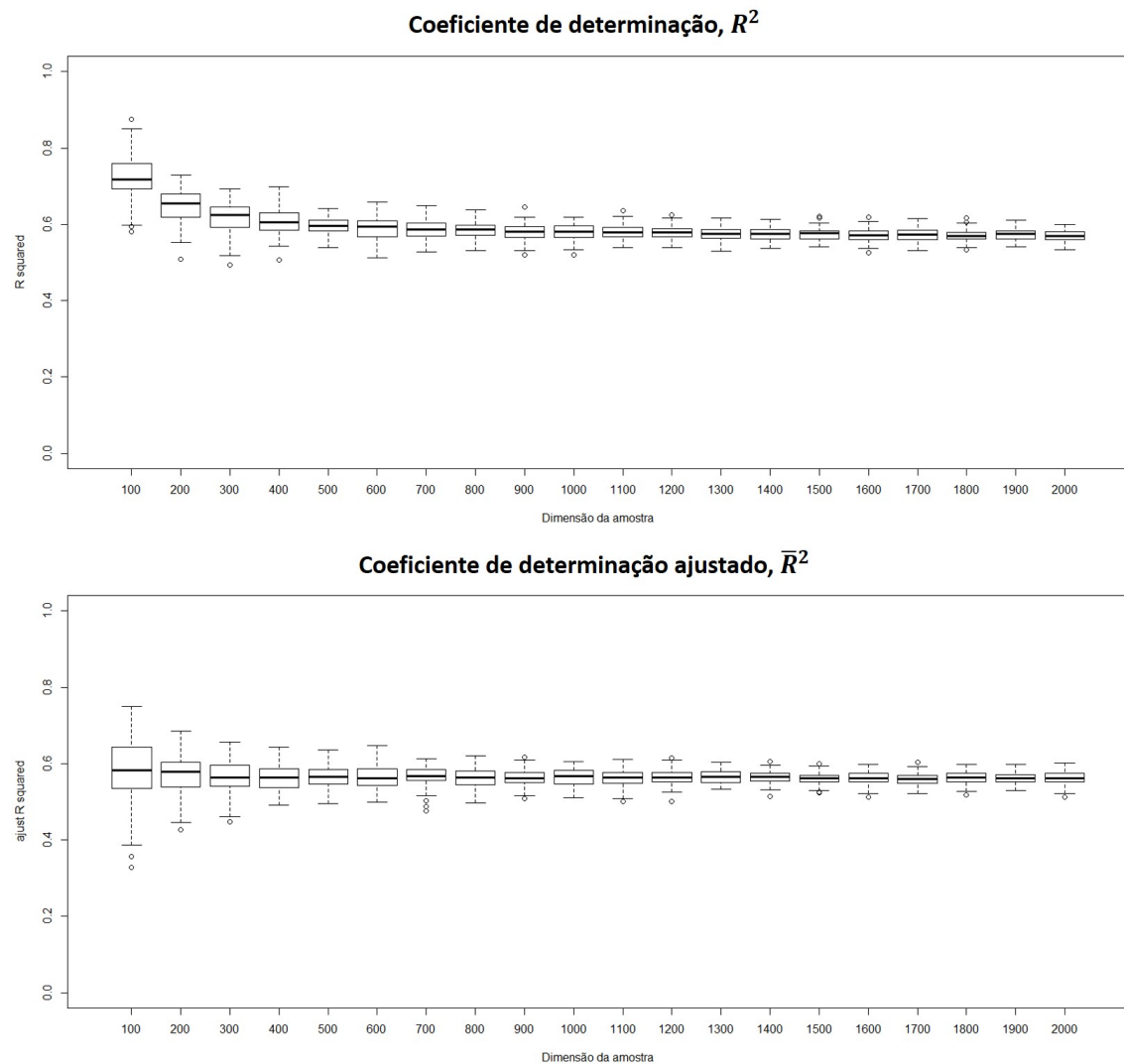


Figura 5.3: Análise do R^2 e do \bar{R}^2

valores verificados para R^2 são bastante mais elevados que os correspondentes valores de \bar{R}^2 . Tal ocorrência deve-se ao facto do coeficiente de determinação ajustado, \bar{R}^2 , penalizar a inclusão de regressores pouco explicativos, enquanto que o coeficiente de determinação, R^2 melhora sempre que se incluem novas variáveis, mesmo que estas tenham pouco poder explicativo.

5.2 Modelo de Regressão Beta

No que diz respeito à aplicação desta metodologia ao modelo de regressão beta, será analisada a variável relativa ao montante do empréstimo em função da dimensão da amostra, à semelhança do efetuado na secção anterior. Será ainda avaliado o impacto da dimensão da amostra no valor do pseudo R^2 e do p -value resultante do teste de razão de verosimilhanças, onde é avaliada a significância global do modelo.

A Figura 5.4 ilustra o comportamento do coeficiente estimado e dos respetivos intervalos de confiança para a variável Montante, à medida que se faz aumentar a amostra.

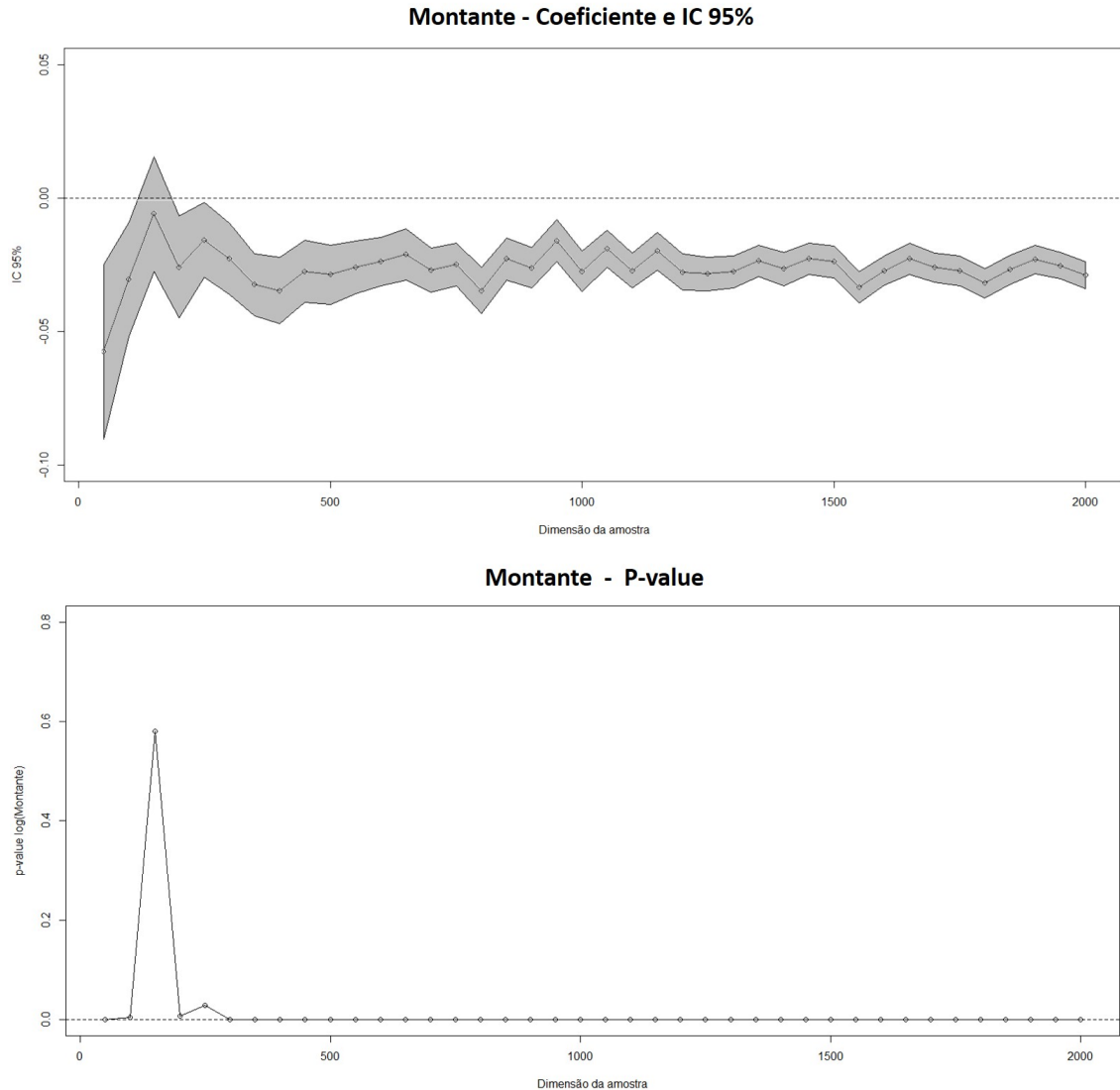


Figura 5.4: Coeficiente, intervalo de confiança a 95% e p -value associados à variável Montante através do modelo de regressão beta

É possível concluir que para pequenas amostras a oscilação do coeficiente estimado é mais elevada comparativamente a amostras de maior dimensão. À semelhança da conclusão retirada no modelo de regressão linear múltipla, parece observar-se que o impacto da dimensão da amostra atinge mais rapidamente o valor do p -value comparativamente ao valor da estimativa do coeficiente de regressão. A partir de amostras de dimensão $n = 300$ a variável Montante torna-se significativa para o modelo, obtendo-se um p -value aproximadamente igual a zero. No conjunto das restantes variáveis, verifica-se que a maioria delas se tornam significativas a 1% a partir de amostras de dimensão $n = 1150$.

CAPÍTULO 5. ANÁLISE DA VARIAÇÃO DOS RESULTADOS EM FUNÇÃO DA DIMENSÃO DAS AMOSTRAS

O gráfico da Figura 5.5 ilustra a evolução dos valores do p -value do teste de razão de verossimilhanças para as várias dimensões da amostra, considerando para cada n , 1000 reamostragens. Observa-se que apenas para amostras de dimensão $n = 100$ existe alguma variabilidade no p -value relativo ao teste de razão de verossimilhança, valores esses que por si só são bastante reduzidos.

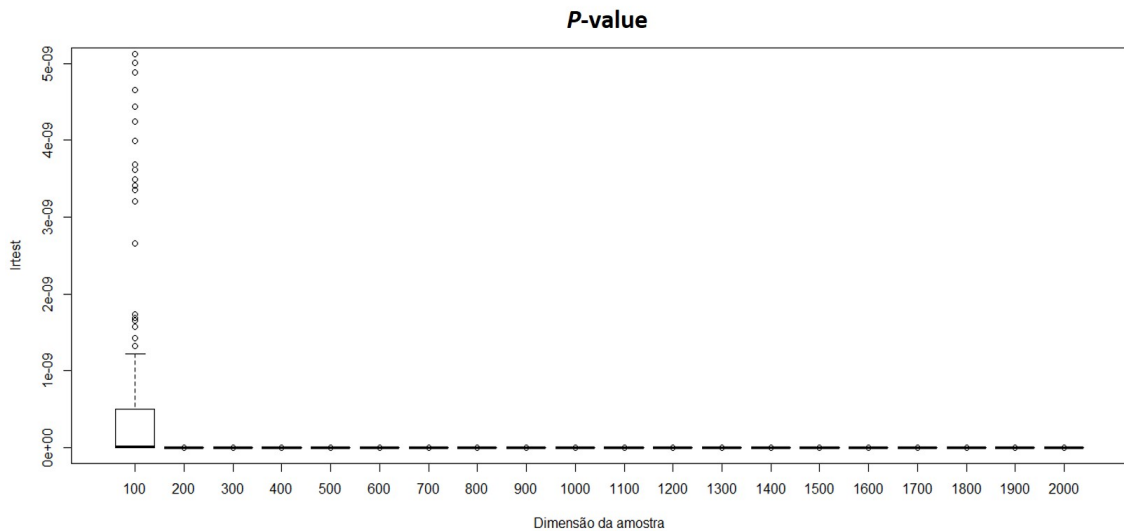


Figura 5.5: Análise do p -value do teste de razão de verossimilhanças

Por fim, a Figura 5.6 representa a evolução do valor do pseudo R^2 para cada uma das diferentes dimensões de amostras consideradas. Verifica-se uma redução dos valores do pseudo R^2 ao longo das amostras de menor dimensão, assim como uma estabilização a partir de amostras de dimensão $n = 500$. Torna-se ainda possível observar uma menor variabilidade do pseudo R^2 quando se consideram amostras de maior dimensão.

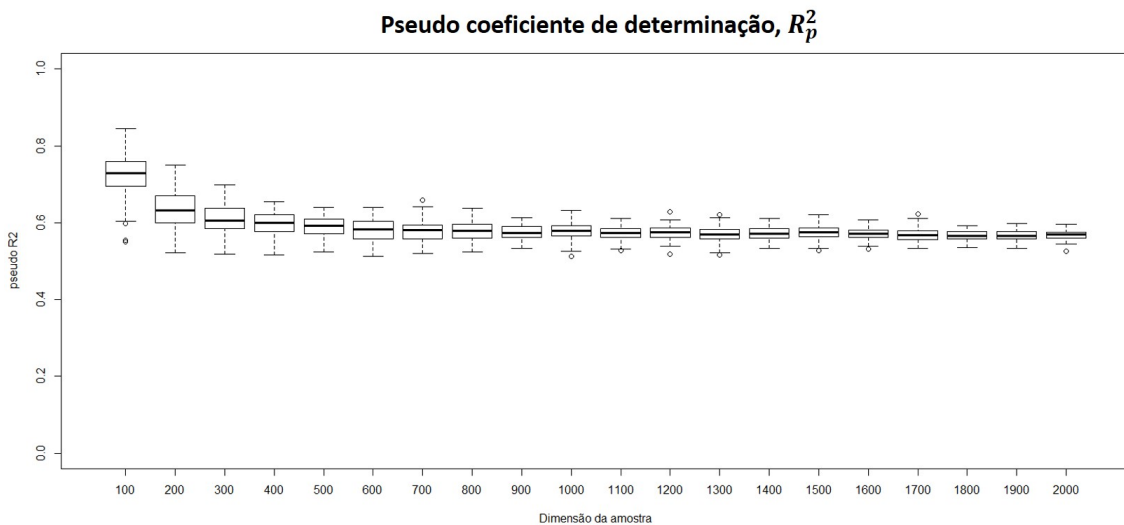


Figura 5.6: Análise do valor do pseudo R^2

A análise efetuada neste capítulo permite verificar que os resultados obtidos por aplicação dos modelos de regressão linear e beta, usando a amostra original, considerada de grande dimensão, também se observam quando são consideradas amostras de menor dimensão. Sendo mesmo possível verificar uma estabilização dos resultados para amostras de dimensão consideravelmente mais pequena.

CONSIDERAÇÕES FINAIS

O presente trabalho só foi possível devido à existência de informação detalhada relativa a todas as novas operações de empréstimos, informação essa proveniente da aplicação do requisito relativo ao reporte da informação individual das taxas de juro.

O objetivo deste trabalho foi, desde início, encontrar um conjunto de fatores determinantes na explicação da fixação das taxas de juro relativas a novas operações de empréstimos concedidos a sociedades não financeiras. Para tal, foram ajustados dois modelos de regressão, o modelo de regressão linear múltipla e o modelo de regressão beta. Os resultados obtidos permitiram compreender a influência que as características consideradas têm nas taxas de juro bancários, sendo possível distinguir os fatores que implicam um agravamento das taxas dos que implica um desagravamento.

Foi possível concluir que empréstimos com montantes e prazos contratuais elevados têm associadas taxas de juro reduzidas. Verificou-se que as novas operações de empréstimos realizadas em instituições financeiras estrangeiras possibilitam, em média, taxas de juro mais reduzidas comparativamente às realizadas em instituições financeiras nacionais. Viu-se também que as taxas praticadas pelos 5 maiores bancos em Portugal apresentam também algumas diferenças no que toca às novas operações de empréstimos. Foi ainda possível concluir que empresas com probabilidades de incumprimento elevadas têm um agravamento das taxas de juro enquanto que empresas de elevado crescimento, exportadoras ou PME líderes, têm um desagravamento das taxas de juro. No que diz respeito aos setores de atividade, verificou-se, por exemplo, um aumento, em média, das taxas de juro relativas a empréstimos concedidos a empresas do setor Imobiliário e da Restauração e uma diminuição das taxas dos setores da Administração Pública, Educação, Saúde e Segurança Social.

Complementarmente à identificação dos fatores que agravam ou desagravam o nível das taxas de juro praticadas, é interessante identificar em que medida esses agravamentos são significativos. Para tal recorreu-se à análise da importância relativa dos regressores onde foi possível identificar a contribuição individual de cada variável para a explicação da variável resposta. Deste modo, concluiu-se que as variáveis respeitantes às características dos empréstimos têm uma maior importância na determinação das taxas de juro, comparativamente às características das empresas a que foram concedidos os empréstimos. No conjunto de características dos empréstimos, as variáveis relativas ao ano, à maturidade, ao montante e à instituição financeira que concedeu o empréstimo são tidas como as mais importantes e as que proporcionam uma maior contribuição para o modelo. Relativamente às características das empresas, as consideradas como mais importantes são a dimensão da empresa, a sua classificação de rating e o número de novas operações realizadas pela empresa no ano em causa. Observou-se que características como o setor de atividade, o facto de a empresa ser exportadora ou de elevado crescimento são relativamente pouco significativas quando comparadas com as características referidas anteriormente.

Uma vez que os resultados do presente trabalho tiveram como base uma amostra de grande dimensão, foi feita uma análise de sensibilidade onde se simularam amostras de menor dimensão com o intuito de avaliar o impacto que a dimensão da amostra tem nos resultados estatísticos obtidos através dos dois modelos de regressão estudados. Verificou-se que, para um nível de significância de 1%, todas as variáveis consideradas se tornam significativas a partir de amostras relativamente pequenas. Pode concluir-se que a significância das variáveis e o ajustamento do modelo não são influenciados pela elevada dimensão da amostra.

É de extrema importância salientar que as conclusões apresentadas são apenas relativas às operações de empréstimos analisadas e poderão não ser representativas da totalidade de novas operações realizadas.

As operações relativas a descobertos bancários não foram tidas em conta no presente trabalho, pelo que, seria interessante, no futuro, abordar essa temática e estudar o seu comportamento, as suas características e quais os fatores determinantes na fixação das respetivas taxas de juro. Seria igualmente interessante analisar os fatores determinantes na fixação das taxas de juro resultantes deste trabalho, comparativamente aos fatores determinantes da fixação das taxas de juro noutros países da União Europeia. Deste modo, seria possível analisar as principais diferenças dos fatores considerados como fundamentais na atribuição das taxas de juro, de país para país. Seria interessante verificar se efetivamente as taxas variam apenas consoante as características do empréstimo ou da empresa, independentemente do país onde se realiza o empréstimo, ou se por outro lado, existe algum tipo de penalização consoante o país tido em consideração.

BIBLIOGRAFIA

- Antunes, P., I. Correia e R. Lameira (2013). “Is micro-data reporting the future for interest rate statistics?” Em: *Session A: Central Credit Registers I* 20, p. 115.
- Batista, J. L. F. (2009). “Verossimilhança e Máxima Verossimilhança”. Em:
- BdP (2009). “Notas Explicativas”. Em: *Boletim Estatístico, Banco de Portugal*.
- Blackwell, M. (2008). “Multiple hypothesis testing: The F-test”. Em: *Matt Blackwell Research*.
- Box, G. E. e D. R. Cox (1964). “An analysis of transformations”. Em: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 211–252.
- Cribari-Neto, F. e A. Zeileis (2009). “Beta regression in R”. Em:
- Delignette-Muller, M. L., C. Dutang et al. (2015). “fitdistrplus: An R package for fitting distributions”. Em: *Journal of Statistical Software* 64(4), pp. 1–34.
- Espinheira, P. L., S. L. Ferrari e F. Cribari-Neto (2008a). “Influence diagnostics in beta regression”. Em: *Computational Statistics & Data Analysis* 52(9), pp. 4417–4431.
- Espinheira, P. L., S. L. Ferrari e F. Cribari-Neto (2008b). “On beta regression residuals”. Em: *Journal of Applied Statistics* 35(4), pp. 407–419.
- Feldman, B. E. (2005). “Relative importance and value”. Em: *Available at SSRN* 2255827.
- Ferrari, S. e F. Cribari-Neto (2004). “Beta regression for modelling rates and proportions”. Em: *Journal of Applied Statistics* 31(7), pp. 799–815.
- Grömping, U. et al. (2006). “Relative importance for linear regression in R: the package relaimpo”. Em: *Journal of statistical software* 17(1), pp. 1–27.
- Hoffman, P. J. (1960). “The paramorphic representation of clinical judgment.” Em: *Psychological bulletin* 57(2), p. 116.
- Johnson, J. W. e J. M. LeBreton (2004). “History and use of relative importance indices in organizational research”. Em: *Organizational Research Methods* 7(3), pp. 238–257.
- Lin, M., H. C. Lucas Jr e G. Shmueli (2013). “Research commentary-too big to fail: large samples and the p-value problem”. Em: *Information Systems Research* 24(4), pp. 906–917.
- Lindeman, R., P. Merenda e R. Gold (1980). *Introduction to bivariate and multivariate analysis*. Scott, Foresman. ISBN: 9780673150998.
- Martinho, R. e A. Antunes (2012a). “Access to credit by non-financial firms”. Em: *Relatório de Estabilidade Financeira, Banco de Portugal*.

- Martinho, R. e A. Antunes (2012b). “Um modelo de scoring para as empresas portuguesas”. Em: *Relatório de Estabilidade Financeira, Banco de Portugal*.
- Osborne, J. W. (2010). “Improving your data transformations: Applying the Box-Cox transformation”. Em: *Practical Assessment, Research & Evaluation* 15(12), pp. 1–9.
- Santos, C. (2013). “Taxas de juro bancárias sobre novas operações de empréstimos concedidos a sociedades não financeiras - Uma primeira apreciação de um novo conjunto de informação microeconómica”. Em: *Relatório de Estabilidade Financeira, Banco de Portugal*.
- Simas, A. B., W. Barreto-Souza e A. V. Rocha (2010). “Improved estimators for a general class of beta regression models”. Em: *Computational Statistics & Data Analysis* 54(2), pp. 348–366.
- Smithson, M. e J. Verkuilen (2006). “A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables.” Em: *Psychological methods* 11(1), p. 54.